

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 42

УТВЕРЖДАЮ
Руководитель образовательной программы
старший преподаватель, к.т.н.

(должность, уч. степень, звание)

В.А. Миклуш

(инициалы, фамилия)



(подпись)

«__18__» июня __2024__ г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Большие данные»
(Наименование дисциплины)

Код направления подготовки/ специальности	09.03.02
Наименование направления подготовки/ специальности	Информационные системы и технологии
Наименование направленности	Информационные технологии в медиаиндустрии
Форма обучения	очная
Год приема	2024

Санкт-Петербург– 2024

Лист согласования рабочей программы дисциплины

Программу составил (а)

доцент, к.т.н., доцент
(должность, уч. степень, звание)


(подпись, дата)

18.06.2024

Н. В. Богословская
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 42
«18» июня 2024 г, протокол № 10/2023-24

Заведующий кафедрой № 42

д.т.н., доц.
(уч. степень, звание)


(подпись, дата)

18.06.2024

С.В. Мичурин
(инициалы, фамилия)

Заместитель директора института №4 по методической работе

доц., к.т.н.
(должность, уч. степень, звание)


(подпись, дата)

18.06.2024

А.А. Фоменкова
(инициалы, фамилия)

Аннотация

Дисциплина «Большие данные» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/ специальности 09.03.02 «Информационные системы и технологии» направленности «Информационные технологии в медиаиндустрии». Дисциплина реализуется кафедрой «№42».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-2 «Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений»

ОПК-1 «Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности»

ОПК-2 «Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности»

ПК-1 «Способен выполнять работы по созданию (модификации) и сопровождению информационных систем»

ПК-7 «Способен проводить аналитическое исследование с применением технологий больших данных, базирующихся в том числе на методах искусственного интеллекта»

Содержание дисциплины охватывает круг вопросов, связанных с основными направлениями в области больших данных: методами консолидации и трансформации больших данных, очистки и предобработки, интеллектуальным (Data Mining) анализом данных (ОПК-1). Big Data (Большие данные) рассматриваются как технология подготовки данных сверхбольшого, непрерывно возрастающего объема, расположенных в распределенных файловых системах и готовых к анализу методами Data Mining и бизнес-аналитики.

При выполнении лабораторных работ используется аналитическая платформа Loginom, внесённая в реестр российских программ (ОПК-2). Loginom является low-code платформой для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации, что позволяет сформировать практические профессиональные навыки анализа, моделирования, прототипирования информационных процессов и систем (ПК-1, ПК-7). Рассматриваются примеры источников, порождающих большие данные: клиентские транзакции розничных торговых сетей, мобильные устройства, банковские данные. Исследуются методы и алгоритмы решения основных задач анализа больших данных: классификация, регрессия, кластеризация, ассоциативные правила (ОПК-1).

В процессе освоения дисциплины продолжается формирование универсальной компетенции УК-2 при выборе оптимальных способов решения поставленных задач с учетом имеющихся ресурсов и ограничений.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, лабораторные работы, самостоятельная работа обучающегося.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы, 144 часа.

Язык обучения по дисциплине «русский».

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины — получение обучающимися необходимых теоретических знаний в области анализа больших данных, формирование умений и навыков в решении прикладных задач бизнес-аналитики в аспекте Big Data.

1.2. Дисциплина входит в состав обязательной части образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Универсальные компетенции	УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.У.1 уметь проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения
Общепрофессиональные компетенции	ОПК-1 Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.В.1 иметь навыки теоретического и экспериментального исследования объектов профессиональной деятельности
Общепрофессиональные компетенции	ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач	ОПК-2.З.1 знать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности ОПК-2.У.1 уметь демонстрировать понимание принципов современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач

	профессиональной деятельности	профессиональной деятельности ОПК-2.В.1 иметь навыки применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности
Профессиональные компетенции	ПК-1 Способен выполнять работы по созданию (модификации) и сопровождению информационных систем	ПК-1.3.1 знать архитектуру, устройство и функционирование информационных систем; сетевые протоколы; основы современных операционных систем; основы современных систем управления базами данных; методы и инструменты для сбора и организации хранения больших данных ПК-1.У.1 уметь разрабатывать документацию для пользователей информационных систем ПК-1.У.3 уметь использовать алгоритмы анализа больших данных и интерпретации полученных результатов
Профессиональные компетенции	ПК-7 Способен проводить аналитическое исследование с применением технологий больших данных, базирующихся в том числе на методах искусственного интеллекта	ПК-7.3.1 знать теоретические и прикладные основы анализа больших данных; типы анализа больших данных, виды аналитики ПК-7.3.2 знать методы интерпретации и визуализации больших данных ПК-7.3.3 знать методы интеллектуального анализа данных ПК-7.У.1 уметь проводить анализ больших данных; осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных, в том числе с применением методов искусственного интеллекта ПК-7.У.2 уметь производить очистку данных для проведения аналитических работ ПК-7.У.3 уметь решать задачи классификации, кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данных ПК-7.В.1 владеть навыками подготовки отчета по результатам аналитических работ с использованием технологий больших данных ПК-7.В.2 владеть приемами разработки и оценки модели больших данных ПК-7.В.3 владеть опытом использования анализа больших данных, в том числе с применением методов искусственного интеллекта

2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

- «Управление данными»,
- «Теория вероятностей и математическая статистика»,
- «Теория информации, данные, знания».

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и могут использоваться при изучении других дисциплин:

- «Методы искусственного интеллекта»,
- «Разработка и анализ требований».

3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№6
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	4/ 144	4/ 144
Из них часов практической подготовки	13	13
Аудиторные занятия, всего час.	51	51
в том числе:		
лекции (Л), (час)	17	17
практические/семинарские занятия (ПЗ), (час)		
лабораторные работы (ЛР), (час)	34	34
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)	36	36
Самостоятельная работа, всего (час)	57	57
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Экз.	Экз.

Примечание: ** кандидатский экзамен

4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ)	ЛР (час)	КП (час)	СРС (час)
Семестр 6					
Раздел 1. Введение в бизнес-аналитику и большие данные					
Тема 1.1. Технологии аналитики данных	3		10		15
Тема 1.2. Методы очистки, предобработки и трансформации					

Раздел 2. Аналитика данных low-code Тема 2.1. Импорт из промышленных источников данных Тема 2.2. Подмодели, проектирование без данных Тема 2.3 Визуализация данных	6		8		15
Раздел 3. Прикладные задачи аналитики больших данных. Введение в алгоритмы Data Mining Тема 3.1 Ассоциативные правила Тема 3.2 Кластеризация Тема 3.3 Классификация Тема 3.4 Регрессия	8		16		27
Итого в семестре:	17		34		57
Итого	17	0	34	0	57

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
1	Раздел 1. Введение в бизнес-аналитику и большие данные Тема 1.1. Технологии аналитики данных Анализ данных. Технологии бизнес-аналитики. Прикладные задачи бизнес-аналитики. Аналитические платформы. Сбор и интеграция данных. Компоненты корпоративной информационной фабрики. Системы управления данными. Преобразование данных. Тема 1.2. Методы очистки, предобработки и трансформации Общие сведения о методах очистки данных. Оценка качества данных. Фильтрация данных. Обработка дубликатов и противоречий. Обнаружение и исключение аномальных значений. Восстановление пропущенных значений. Подготовка данных в аналитических системах. Трансформация данных в аналитических системах. Компоненты «Группировка», «Дата и время», «Квантование», «Скользящее окно», «Кросс-таблица». Основы регулярных выражений. Синтаксис регулярных выражений Perl.
2	Раздел 2. Аналитика данных low-code Тема 2.1. Импорт из промышленных источников данных Узел импорта Excel-файл. Настройка импорта. Использование подключения в сценарии. Узел импорта База данных. Визуальный интерфейс импорта. Импорт с помощью SQL. Переменные в SQL-запросе. Чтение переменных из БД. Тема 2.2. Подмодели, проектирование без данных Производные компоненты. Производные узлы. Создание производного компонента. Наследование в Logiот. Обновление и восстановление конфигурации производного узла. Переопределение производных узлов. Визуализаторы в производных компонентах. Проектирование сверху вниз. Проектирование без данных. Пример: проектирование подмодели для расчета скидки. Ограничения проектирования без данных.

	Тема 2.3. Визуализация данных Визуализация и бизнес-аналитика. Цели и задачи визуализации. Этапы визуализации. Методы визуализации. Визуализатор Статистика. Визуализаторы общего назначения. Таблицы. Графики. Диаграммы. Круговые диаграммы. Гистограммы. Визуализатор "Диаграмма". Сложные визуализаторы общего назначения. OLAP-анализ. Географические карты. Тепловые карты. Плоское дерево. Диаграмма связей. Облако тегов. Пузырьковая диаграмма. Диаграмма рассеяния. Визуализатор OLAP-куб.
3	Раздел 3. Прикладные задачи аналитики больших данных. Введение в алгоритмы Data Mining Тема 3.1. Алгоритм поиска ассоциативных правил. Тема 3.2. Кластеризация данных. Тема 3.3. Классификация. Тема 3.4. Регрессия.

4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Учебным планом не предусмотрено					
Всего					

4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Семестр 6				
1	Основы работы в Logiplot. Сценарии. Экспорт и импорт данных. Настройка портов. Компоненты Калькулятор, Фильтр строк, Сортировка.	2		1
2	Структуризация сценария – Подмодели. Проектирование без данных: переменные и параметризация. Компоненты Группировка, Условие, Замена.	2		2
3	Обогащение данных: Слияние, Соединение, Дополнение, Объединение. Компоненты Дата и время, Таблица в переменные, Квантование.	2		1
4	Визуализация данных. Аналитическая	2		2

	отчетность.			
5	Формальный язык поиска и манипуляций в больших массивах данных: регулярные выражения.	2		1
6	Трансформация данных. Практическая задача 1 - Балльная скоринговая карта.	2	1	1
7	Производный компонент многократного использования. Практическая задача 2 – Анализ чеков в розничной сети.	2	1	2
8	Очистка и предобработка. Практическая задача 3 – Очистка и нормализация.	2	1	1
9	Консолидация данных. Импорт данных из промышленных источников.	2	2	2
10	Data Mining: Ассоциативные правила. Генерация правил, интерпретация, визуализация результатов.	4	2	3
11	Data Mining: Кластеризация. Практическая задача 4 - Сегментация клиентов телекоммуникационной компании.	4	2	3
12	Data Mining: Классификация. Практическая задача 5 – ABC и XYZ анализ	4	2	3
13	Data Mining: Регрессия. Практическая задача 6 - Прогнозирование продажи объекта недвижимости на основе логистической регрессии.	4	2	3
Всего		34	13	

4.5. Курсовое проектирование/ выполнение курсовой работы
Учебным планом не предусмотрено

4.6. Самостоятельная работа обучающихся
Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 6, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	20	20
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)	17	17
Домашнее задание (ДЗ)		
Контрольные работы заочников (КРЗ)		
Подготовка к промежуточной аттестации (ПА)	20	20
Всего:	57	57

5. Перечень учебно-методического обеспечения
для самостоятельной работы обучающихся по дисциплине (модулю)
Учебно-методические материалы для самостоятельной работы обучающихся указаны в
п.п. 7-11.

6. Перечень печатных и электронных учебных изданий
Перечень печатных и электронных учебных изданий приведен в таблице 8.
Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
519.6/.8(ГУАП) А 66	Андронов, Сергей Александрович. Модели и методы в системах поддержки принятия решений : учебное пособие / С. А. Андронов ; С.-Петербург. гос. ун-т аэрокосм. приборостроения. - СПб. : Изд-во ГУАП, 2008. - 176 с. : рис., табл. - Библиогр.: с. 173 - 175 (54 назв.). - ISBN 978-5-8088-0374-9 : 70.00 р. - Текст : непосредственный.	119
https://znanium.com/catalog/ product/2051477	Григорьев, А. А. Передача, хранение и обработка больших объемов научных данных : учебное пособие / А.А. Григорьев, Е.А. Исаев, П.А. Тарасов. — Москва : ИНФРА-М, 2024. — 207 с. — (Высшее образование). — DOI 10.12737/1073525. - ISBN 978-5-16- 018850-8. - Текст : электронный. - URL: https://znanium.com/catalog/product/2051477 (дата обращения: 26.08.2023). — Режим доступа: по подписке.	
https://znanium.com/catalog/ product/1027845	Лесковец, Ю. Анализ больших наборов данных / Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман ; пер. с англ. А.А.Слинкина. - Москва : ДМК Пресс, 2016. - 498 с. - ISBN 978-5-97060-190-7. - Текст : электронный. - URL:	

	https://znanium.com/catalog/product/1027845 (дата обращения: 26.08.2023). – Режим доступа: по подписке.	
https://znanium.com/catalog/product/1232190	Адлер, Ю. П. Статистическое управление процессами. «Большие данные» : учебное пособие / Ю. П. Адлер, Е. А. Черных. - Москва : Изд. Дом МИСиС, 2016. - 52 с. - ISBN 978-5-87623-969-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/1232190 (дата обращения: 26.08.2023). – Режим доступа: по подписке.	

7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
https://help.loginom.ru/userguide/#loginom-%E2%80%94-rukovodstvo-polzovatelya	Loginom – руководство пользователя

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
1	Аналитическая платформа Loginom. Бесплатная версия для некоммерческого использования Community https://loginom.ru/

8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
1	Энциклопедия по бизнес-анализу https://wiki.loginom.ru/?_ga=2.204220440.988103975.1592928053-409674285.1589973143
2	Общедоступные наборы данных Datasets https://www.kaggle.com/

9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Мультимедийная лекционная аудитория	
2	Компьютерный класс	

10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Экзамен	Список вопросов к экзамену; Экзаменационные билеты; Тесты.

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 – Критерии оценки уровня сформированности компетенций

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«отлично» «зачтено»	– обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«хорошо» «зачтено»	– обучающийся твердо усвоил программный материал, грамотно и по существу, излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения; – делает выводы и обобщения; – владеет системой специализированных понятий.
«удовлетворительно» «зачтено»	– обучающийся усвоил только основной программный материал, по существу, излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления;

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
	– слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий.
«неудовлетворительно» «не зачтено»	– обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний; – не может аргументировать научные положения; – не формулирует выводов и обобщений.

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена	Код индикатора
1	Технологии бизнес-аналитики и большие данные	УК-2.У.1
2	Жизненный цикл аналитических проектов Big Data. Интеграция, подготовка, очистка, анализ, интерпретация. Особенности задач с Big Data.	ОПК-1.В.1
3	Аналитические платформы	ОПК-2.3.1
4	Технологии KDD и Data Mining	ОПК-2.У.1
5	Технологии подготовки данных к анализу	ОПК-2.В.1
6	Подготовка данных к анализу. Особенности данных, накопленных в компаниях. Формализация данных. Методы сбора данных.	ПК-1.3.1
7	Отчетность и визуализация данных в аналитических платформах	ПК-1.У.1
8	Ассоциативные правила. Назначение. Примеры использования. Базовые понятия. Методы поиска ассоциативных правил. Метрики ассоциативных правил. Интерпретация ассоциативных правил.	ПК-1.У.3
9	Требования к данным. Формы представления данных. Типы шкал. Информативность данных. Представление наборов данных. Особенности больших данных.	ПК-7.3.1
10	Визуализация данных. Визуализаторы общего назначения и специализированные. OLAP-анализ. Визуализаторы оценки качества моделирования. Визуализаторы, интерпретации результатов анализа. Концепция анализа данных Дж. Тьюки.	ПК-7.3.2
11	Кластерный анализ. Формальное описание алгоритма кластеризации. Типы кластерных структур. Меры сходства. Методы определения расстояния между кластерами. Иерархические методы кластеризации. Алгоритмы k-means и g-means. EM- кластеризация. Кластеризация категориальных данных. Алгоритм CLOPE.	ПК-7.3.3
12	Ансамбли моделей в аналитике. Пример совместного использования ABC и XYZ моделей анализа.	ПК-7.У.1

13	Очистка данных. Оценка качества данных. Методы оценки качества данных. Фильтрация данных, обработка дубликатов и противоречий. Выявление аномальных значений. Сокращение размерности данных. Сокращение числа признаков и записей.	ПК-7.У.2
14	Модели и задачи Data Mining. Современное развитие классификации задач Data Mining. Основные свойства и характеристик методов Data Mining.	ПК-7.У.3
15	Аналитическая отчетность. Роль визуализации в аналитике данных. Методика извлечения и обогащения данных.	ПК-7.В.1
16	Требования к данным. Формы представления данных. Типы шкал. Информативность данных. Представление наборов данных. Особенности больших данных.	ПК-7.В.2
17	Аналитическая пирамида аналитики. Инструменты аналитики данных. Современное понятие анализа данных. Особенности задач нового типа.	ПК-7.В.3

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета	Код индикатора
	Учебным планом не предусмотрено	

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
	Учебным планом не предусмотрено

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов	Код индикатора
1	<i>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</i> Одной из проблем правовой охраны big data является следующая особенность – 1 вторичность и возможность многократного использования big data для достижения разнообразных целей. 2 возможность многократного использования big data для достижения разнообразных целей. 3 вторичность использования big data для достижения разнообразных целей. 4 вторичность и невозможность многократного использования big data для достижения разнообразных целей.	УК-2
2	<i>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</i> Регулирование персональных данных, которые могут оказаться	УК-2

	частью big data, в России осуществляется с применением ряда нормативно-правовых актов: 1. № 152-ФЗ «О персональных данных»; 2. № 149-ФЗ «Об информации, информационных технологиях и о защите информации»; 3. ТК РФ Статья 86. Общие требования при обработке персональных данных работника и гарантии их защиты; 4. № 152-ФЗ «О личных данных клиентов и покупателей»; 5. № 125-ФЗ «Об архивном деле в Российской Федерации».													
3	Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце. Big data можно описать как непрерывающийся поток данных, поступающий из различных источников, которые можно условно объединить в две основные группы: <table><tr><td>1. Интернет вещей (англ. Internet of Things, IoT, включающий в свой состав и промышленный интернет вещей (англ. Industrial Internet of Things, IIoT)), который «поставляет» информацию со всевозможных действующих датчиков, котроллеров, приборов учета потребления — это</td><td>А. технические источники</td></tr><tr><td>2. Искусственный интеллект (англ. Artificial Intelligence, AI) — это</td><td>В. социальные источники</td></tr><tr><td>3. Машинное обучение (англ. Machine Learning, ML) — это</td><td></td></tr><tr><td>4. Социальные медиа (англ. social media) — это</td><td></td></tr><tr><td>5. Розничную торговлю (англ. retail) — это</td><td></td></tr><tr><td>6. Здравоохранение — это</td><td></td></tr></table>	1. Интернет вещей (англ. Internet of Things, IoT, включающий в свой состав и промышленный интернет вещей (англ. Industrial Internet of Things, IIoT)), который «поставляет» информацию со всевозможных действующих датчиков, котроллеров, приборов учета потребления — это	А. технические источники	2. Искусственный интеллект (англ. Artificial Intelligence, AI) — это	В. социальные источники	3. Машинное обучение (англ. Machine Learning, ML) — это		4. Социальные медиа (англ. social media) — это		5. Розничную торговлю (англ. retail) — это		6. Здравоохранение — это		УК-2
1. Интернет вещей (англ. Internet of Things, IoT, включающий в свой состав и промышленный интернет вещей (англ. Industrial Internet of Things, IIoT)), который «поставляет» информацию со всевозможных действующих датчиков, котроллеров, приборов учета потребления — это	А. технические источники													
2. Искусственный интеллект (англ. Artificial Intelligence, AI) — это	В. социальные источники													
3. Машинное обучение (англ. Machine Learning, ML) — это														
4. Социальные медиа (англ. social media) — это														
5. Розничную торговлю (англ. retail) — это														
6. Здравоохранение — это														
4	Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо. Этапы построения ML-модели можно описать следующей схемой: А. Подготовка и очистка данных. В. Конструирование признаков. С. Разбиение на множества. Д. Кодирование и стандартизация признаков. Е. Построение модели. Ф. Оценка качества. Г. Эксплуатация модели.	УК-2												
5	Прочитайте текст и запишите развернутый обоснованный ответ. Обработка больших данных часто включает конфиденциальные или чувствительные данные. Какие правовые вопросы должны быть учтены при работе с большими данными в этом аспекте?	УК-2												
6	Прочитайте текст, выберите правильный ответ и запишите	ОПК-1												

	<i>аргументы, обосновывающие выбор ответа.</i> В чем состоит алгоритм k-средних? 1. из исходного множества случайным образом выбираются k-центров кластеров; 2. рассчитывается диаметр множества, делится на k, и кругами, равными полученному значению, покрывается все множество; внутри каждого круга находится центр кластера; 3. множество делится на части с помощью k -плоскостей; 4. множество объектов делится на k частей.							
7	<i>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</i> Перечислите четыре основные характеристики Big Data: 1. Virtualization 2. Volume 3. Velocity 4. Variety 5. Vehicle 6. Value	ОПК-1						
8	<i>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце.</i> Направления работы по управлению Big Data должны основываться на определенных принципах. <table><tr><td>1. Горизонтальная масштабируемость.</td><td>А. Любая система, которая подразумевает обработку больших данных, должна быть расширяемой.</td></tr><tr><td>2. Отказоустойчивость.</td><td>В. Методы работы с большими данными должны учитывать такие возможности и переживать их без каких-либо значимых последствий.</td></tr><tr><td>3. Локальность данных.</td><td>С. По возможности обрабатывать данные на той же машине, где они хранятся.</td></tr></table>	1. Горизонтальная масштабируемость.	А. Любая система, которая подразумевает обработку больших данных, должна быть расширяемой.	2. Отказоустойчивость.	В. Методы работы с большими данными должны учитывать такие возможности и переживать их без каких-либо значимых последствий.	3. Локальность данных.	С. По возможности обрабатывать данные на той же машине, где они хранятся.	ОПК-1
1. Горизонтальная масштабируемость.	А. Любая система, которая подразумевает обработку больших данных, должна быть расширяемой.							
2. Отказоустойчивость.	В. Методы работы с большими данными должны учитывать такие возможности и переживать их без каких-либо значимых последствий.							
3. Локальность данных.	С. По возможности обрабатывать данные на той же машине, где они хранятся.							
9	<i>Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо.</i> Перечислите шкалы в порядке возрастания количества измеряемых ими характеристик: А. Номинальная шкала. В. Порядковая шкала. С. Интервальная шкала. D. Шкала отношений.	ОПК-1						
10	<i>Прочитайте текст и запишите развернутый обоснованный ответ.</i> Числовые данные являются дискретными если имеют	ОПК-1						
11	<i>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</i> К методам сбора big data относятся: 1.получение из учетных систем, получение из косвенных источников, использование открытых источников, проведение	ОПК-2						

	собственных мероприятий по сбору данных, ввод данных вручную. 2. получение из учетных систем, использование открытых источников, ввод данных вручную. 3. получение из учетных систем, получение из косвенных источников, использование открытых источников, проведение собственных мероприятий по сбору данных. 4. получение из учетных систем, получение из косвенных источников, проведение собственных мероприятий по сбору данных, ввод данных вручную					
12	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов. Технология Knowledge Discovery in Databases (KDD) обнаружение знаний в базах данных – это последовательность действий по 1. подготовке данных, 2. выбору информативных признаков, 3. очистке, 4. построение моделей, 5. эксплуатации моделей, 5. интерпретации полученных результатов.	ОПК-2				
13	Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце. <table><tr><td>1. Непрерывные данные — это</td><td>А. данные, которые могут принимать любые значения в некотором интервале. Над ними можно производить арифметические операции, и они имеют смысл.</td></tr><tr><td>2. Дискретные данные — это</td><td>В. данные, которые являются значениями признака. С такими данными не могут быть произведены никакие арифметические действия (не имеют смысла).</td></tr></table>	1. Непрерывные данные — это	А. данные, которые могут принимать любые значения в некотором интервале. Над ними можно производить арифметические операции, и они имеют смысл.	2. Дискретные данные — это	В. данные, которые являются значениями признака. С такими данными не могут быть произведены никакие арифметические действия (не имеют смысла).	ОПК-2
1. Непрерывные данные — это	А. данные, которые могут принимать любые значения в некотором интервале. Над ними можно производить арифметические операции, и они имеют смысл.					
2. Дискретные данные — это	В. данные, которые являются значениями признака. С такими данными не могут быть произведены никакие арифметические действия (не имеют смысла).					
14	Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо. Если в сжатой форме описывать процесс сбора и обработки большого массива данных, то стоит выделить основные этапы: А. постановка задачи для аналитической программы; В. программа осуществляет сбор данных с их параллельной подготовкой (декодировка, отсев мусора, удаление нерелевантной информации); С. выбор алгоритма анализа данных; Д. обучение программы выбранному алгоритму с дальнейшим анализом обнаруженных закономерностей.	ОПК-2				
15	Прочитайте текст и запишите развернутый обоснованный ответ. В большинстве случаев полученные необработанные данные хранятся в так называемом «озере данных» — Data Lake. Какие форматы и уровни структуризации информации при этом возможны?	ОПК-2				
16	Прочитайте текст, выберите правильный ответ и запишите	ПК-1				

	<i>аргументы, обосновывающие выбор ответа.</i> Базовые понятия теории ассоциативных правил: 1. Предметный набор, транзакция. 2. Ассоциация, корреляция, взаимодействие. 3. Взаимозависимость, условие, следствие. 4. Структура, элемент, зависимость.											
17	<i>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</i> Существует различные типы аналитики big data: 1. описательная, 2. диагностическая, 3. предиктивная, 4. предписывающая, 5. генеративная.	ПК-1										
18	<i>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце.</i> <table><tr><td>1. Кластеризация – это</td><td>А. описательные (англ. descriptive) методы обнаружения знаний</td></tr><tr><td>2. Ассоциативные правила – это</td><td>В. предсказательные (англ. predictive) методы обнаружения знаний</td></tr><tr><td>3. Сокращение описания – это</td><td></td></tr><tr><td>4.Классификация – это</td><td></td></tr><tr><td>5.Регрессия – это</td><td></td></tr></table>	1. Кластеризация – это	А. описательные (англ. descriptive) методы обнаружения знаний	2. Ассоциативные правила – это	В. предсказательные (англ. predictive) методы обнаружения знаний	3. Сокращение описания – это		4.Классификация – это		5.Регрессия – это		ПК-1
1. Кластеризация – это	А. описательные (англ. descriptive) методы обнаружения знаний											
2. Ассоциативные правила – это	В. предсказательные (англ. predictive) методы обнаружения знаний											
3. Сокращение описания – это												
4.Классификация – это												
5.Регрессия – это												
19	<i>Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо.</i> Этапы кластеризации: А. Формулировка проблемы; В. Выбор способа измерения расстояния; С. Выбор метода кластеризации; Д. Принятие решения о количестве кластеров; Е. Оценка качества кластеризации; Ф. Интерпретация и профилирование кластеров.	ПК-1										
20	<i>Прочитайте текст и запишите развернутый обоснованный ответ.</i> Data Mining— собирательное название, используемое для обозначения совокупности методов. Data Mining — обнаружение в «сырых» данных ранее.....	ПК-1										
21	<i>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа.</i> 1. Наиболее распространен среди неиерархических методов алгоритм k-средних, также называемый быстрым кластерным анализом. 2. Наиболее распространен среди иерархических методов алгоритм k-средних, также называемый быстрым кластерным анализом. 3. Наиболее распространен среди неиерархических методов алгоритм g-средних, также называемый быстрым кластерным анализом. 4. Наиболее распространен среди иерархических методов	ПК-7										

	алгоритм g-средних, также называемый быстрым кластерным анализом.					
22	Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов. 1. Наиболее распространен среди неиерархических методов алгоритм k-средних, также называемый быстрым кластерным анализом. 2. В отличие от иерархических методов, которые не требуют предварительных предположений относительно числа кластеров, для метода k-means необходимо иметь гипотезу о наиболее вероятном количестве кластеров. 3. Алгоритм k-средних строит k кластеров, расположенных на возможно больших расстояниях друг от друга. 4. Выбор числа k может базироваться на результатах предшествующих исследований, теоретических соображениях или интуиции. 5. Алгоритм k-средних строит k кластеров, расположенных как можно ближе друг к другу.	ПК-7				
23	Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце. <table><tr><td>1. G-means - популярный алгоритм кластеризации</td><td>А. с автоматическим выбором числа кластеров.</td></tr><tr><td>2. K-means - популярный алгоритм кластеризации</td><td>В. необходимо иметь гипотезу о наиболее вероятном количестве кластеров.</td></tr></table>	1. G-means - популярный алгоритм кластеризации	А. с автоматическим выбором числа кластеров.	2. K-means - популярный алгоритм кластеризации	В. необходимо иметь гипотезу о наиболее вероятном количестве кластеров.	ПК-7
1. G-means - популярный алгоритм кластеризации	А. с автоматическим выбором числа кластеров.					
2. K-means - популярный алгоритм кластеризации	В. необходимо иметь гипотезу о наиболее вероятном количестве кластеров.					
24	Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо. В технологии Bining процедура Classing определяет, как будет представлена переменная с точки зрения ее уникальных значений. Процедура состоит из двух шагов: А. формирование Fine Classing, В. формирование Coarse Classing.	ПК-7				
25	Прочитайте текст и запишите развернутый обоснованный ответ. «Модель» в Python – это объект класса, который имеет....	ПК-7				

Примечание:

Система оценивания тестовых заданий:

1 тип) Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора считается верным, если правильно указана цифра и приведены конкретные аргументы, используемые при выборе ответа. Полное совпадение с верным ответом оценивается 1 баллом, неверный ответ или его отсутствие – 0 баллов.

2 тип) Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора считается верным, если правильно указаны цифры и приведены конкретные аргументы, используемые при выборе ответов. Полное совпадение с верным ответом оценивается 1 баллом, если допущены ошибки или ответ отсутствует – 0 баллов.

3 тип) Задание закрытого типа на установление соответствия считается верным, если установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого столбца). Полное совпадение с верным ответом оценивается 1 баллом, неверный ответ или его отсутствие – 0 баллов.

4 тип) Задание закрытого типа на установление последовательности считается верным, если правильно указана вся последовательность цифр. Полное совпадение с верным ответом оценивается 1 баллом, если допущены ошибки или ответ отсутствует – 0 баллов.

5 тип) Задание открытого типа с развернутым ответом считается верным, если ответ совпадает с эталонным по содержанию и полноте. Правильный ответ за задание оценивается в 3 балла, если допущена одна ошибка

\ неточность \ ответ правильный, но не полный - 1 балл, если допущено более 1 ошибки \ ответ неправильный \ ответ отсутствует – 0 баллов.

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
	Не предусмотрено

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

- Вступительная часть определяет название темы, план и цель лекции;
- В основной части реализуется содержание темы, все главные узловые вопросы, рассматриваются приемы и технологии текущей темы с применением средств визуализации. Каждый вопрос заканчивается краткими выводами, логически подводящими студентов к следующему вопросу лекции;
- Заключительная часть обобщает в кратких формулировках основные идеи лекции, логически завершая ее как целостное рассмотрение темы.

11.2. Методические указания для обучающихся по выполнению лабораторных работ

В ходе выполнения лабораторных работ обучающийся должен углубить и закрепить знания, практические навыки, овладеть современной методикой и техникой эксперимента в соответствии с квалификационной характеристикой обучающегося. Выполнение лабораторных работ состоит из экспериментально-практической, расчетно-аналитической частей и контрольных мероприятий.

Выполнение лабораторных работ обучающимся является неотъемлемой частью изучения дисциплины, определяемой учебным планом, и относится к средствам, обеспечивающим решение следующих основных задач обучающегося:

- приобретение навыков исследования процессов, явлений и объектов, изучаемых в рамках данной дисциплины;
- закрепление, развитие и детализация теоретических знаний, полученных на лекциях;
- получение новой информации по изучаемой дисциплине;
- приобретение навыков самостоятельной работы с лабораторным оборудованием и приборами.

Задание и требования к проведению лабораторных работ

Задание на лабораторную работу выкладывается в личный кабинет студента. Обязательными мероприятиями являются: установка указанного программного обеспечения, знакомство со справочными материалами. Лабораторная работа выполняется на компьютерах в аудиториях кафедры или на личных ноутбуках.

Структура и форма отчета о лабораторной работе

Отчет о выполнении лабораторной работы должен включать следующие позиции: постановку задачи, описание исследуемой предметной области, пошаговое описание технологии выполнения с необходимыми комментариями к разработанному сценарию обработки данных, результаты и выводы.

Требования к оформлению отчета о лабораторной работе

Изложение текста и оформление работы следует выполнять в соответствии с требованиями, представленными на сайте ГУАП <https://guap.ru/regdocs/docs/uch>

11.3. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся являются:

- учебно-методический материал по дисциплине.

11.4. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Учитываются следующие позиции: посещение лекций, оценки, полученные по результатам защиты лабораторных работ, посещаемость. Все перечисленные показатели оцениваются в баллах, суммируются и являются базовой оценкой при проведении промежуточной аттестации.

11.5. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя:

- экзамен – форма оценки знаний, полученных обучающимся в процессе изучения всей дисциплины или ее части, навыков самостоятельной работы, способности применять их для решения практических задач. Экзамен, как правило, проводится в период экзаменационной сессии и завершается аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой