

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

Кафедра № 2

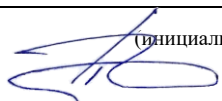
УТВЕРЖДАЮ

Руководитель образовательной программы

проф., д.т.н., проф.

(должность, уч. степень, звание)

Е.А. Перепелкин


(инициалы, фамилия)

(подпись)

«09» февраля 2026 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Интеллектуальный анализ больших данных»
(Наименование дисциплины)

Код направления подготовки/ специальности	09.04.03
Наименование направления подготовки/ специальности	Прикладная информатика
Наименование направленности/ специализации	Цифровая аналитика производственных систем
Форма обучения	очная
Год приема	2026

Лист согласования рабочей программы дисциплины

Программу составил (а)

проф., д.ф.-м.н., доц.
(должность, уч. степень, звание)



06.02.2026

(подпись, дата)

А.А. Макаров

(инициалы, фамилия)

Программа одобрена на заседании кафедры № 2

«06» февраля 2026 г, протокол № 7/25-26

Заведующий кафедрой № 2

д.ф.-м.н., проф.
(уч. степень, звание)



06.02.2026

(подпись, дата)

В.Г. Фарафонов

(инициалы, фамилия)

Заместитель директора института ФПТИ по методической работе

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



06.02.2026

(подпись, дата)

Н.Ю. Ефремов

(инициалы, фамилия)

Аннотация

Дисциплина «Интеллектуальный анализ больших данных» входит в образовательную программу высшего образования – программу магистратуры по направлению подготовки/ специальности 09.04.03 «Прикладная информатика» направленности/специализации «Цифровая аналитика производственных систем». Дисциплина реализуется кафедрой «№2».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-1 «Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий»

ПК-2 «Способность проектировать архитектуру информационной системы»

Содержание дисциплины охватывает круг вопросов, связанных с интеллектуальным анализом больших данных.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия, самостоятельная работа обучающегося.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме экзамена (2 семестр).

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц, 180 часов.

Язык обучения по дисциплине «русский»

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины

формирование системы понятий, знаний, умений и навыков в области интеллектуального анализа цифровых данных.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, 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-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.В.2 владеть навыками использования алгоритмов и цифровых средств, предназначенных для анализа информации и данных
Профессиональные компетенции	ПК-2 Способность проектировать архитектуру информационной системы	ПК-2.3.1 знать основы теории систем и системного анализа

2. Место дисциплины в структуре ОП

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

- «Специальные разделы высшей математики»,
- «Современные технологии разработки ПО».

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

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
		№2
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	5/ 180	5/ 180
Из них часов практической подготовки	9	9

Аудиторные занятия , всего час.	34	34
в том числе:		
лекции (Л), (час)	17	17
практические/семинарские занятия (ПЗ), (час)	17	17
лабораторные работы (ЛР), (час)		
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)	54	54
Самостоятельная работа , всего (час)	92	92
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.)	Экз.,	Экз.,

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) (час)	ЛР (час)	КП/КР (час)	СР (час)
Семестр 2					
Раздел 1. Методы обработки сигналов и изображений	17	17			92
Итого в семестре:	17	17			92
Итого	17	17	0	0	92

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	<p>Тема 1. Введение в большие данные и интеллектуальный анализ.</p> <p>Определение и характеристики больших данных (модель «V»: Volume, Velocity, Variety, Veracity, Value). Понятие интеллектуального анализа данных (Data Mining). Место KDD-процесса в анализе данных. Основные этапы анализа: от постановки задачи до интерпретации результатов.</p> <p>Области применения: бизнес-аналитика, биоинформатика, интернет вещей, социальные сети. Обзор современных инструментов и платформ (Hadoop, Spark, облачные решения).</p> <p>Тема 2. Хранение и управление большими данными.</p> <p>Распределённые файловые системы (HDFS, Amazon S3). Модель MapReduce: принципы параллельной обработки.</p> <p>NoSQL базы данных: типы (документоориентированные, колоночные, графовые, ключ-значение) и примеры (MongoDB, Cassandra, Neo4j). Озёра данных (Data Lakes) и</p>

	<p>хранилища данных (Data Warehouses). Концепция Data Mesh и управление качеством данных.</p> <p>Тема 3. Предобработка и очистка данных. Основные проблемы «грязных» данных: пропуски, выбросы, дубликаты, несогласованность. Методы обработки пропущенных значений: удаление, импутация (средним, медианой, с помощью моделей). Нормализация и стандартизация признаков (Min-Max, Z-score). Кодирование категориальных признаков (One-Hot Encoding, Label Encoding). Отбор признаков: фильтрационные, обёрточные и встроенные методы. Снижение размерности (PCA).</p>
2	<p>Тема 4. Методы интеллектуального анализа данных. Постановка задач обучения с учителем: классификация и регрессия. Деревья решений: алгоритмы построения (ID3, C4.5, CART), критерии разбиения (энтропия, индекс Джини). Ансамблевые методы: случайный лес (Random Forest), градиентный бустинг (XGBoost, LightGBM, CatBoost). Метод опорных векторов (SVM) и логистическая регрессия. Оценка качества моделей: accuracy, precision, recall, F1-мера, ROC-AUC, MSE, R². Кросс-валидация.</p> <p>Тема 5. Кластерный анализ и поиск аномалий. Постановка задачи кластеризации: обучение без учителя. Алгоритм K-means: принцип работы, выбор числа кластеров (метод локтя, силуэтный коэффициент). Иерархическая кластеризация: агломеративный и дивизимный подходы, дендрограммы. Плотностные методы: DBSCAN, OPTICS — преимущества для данных произвольной формы. Обнаружение аномалий (Anomaly Detection): статистические методы, изолирующий лес (Isolation Forest), one-class SVM.</p> <p>Тема 6. Анализ ассоциативных правил и рекомендательные системы. Задача поиска частых наборов и ассоциативных правил. Метрики: support, confidence, lift. Алгоритм Apriori и свойство антимонотонности. Алгоритм FP-Growth: построение дерева частых шаблонов. Рекомендательные системы: контентная фильтрация и коллаборативная фильтрация. Матричные разложения (SVD, ALS) и гибридные подходы. Проблема холодного старта.</p>

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
Семестр 2					
1	Первичный анализ данных, выявление закономерностей и подготовка данных	ситуационная задача	9		1

	к моделированию			
2	Классификация и кластеризация, оценка качества построенных моделей	ситуационная задача	8	1
Всего			17	

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
Учебным планом не предусмотрено				
Всего				

4.5. Выполнение курсового проекта/ 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 2, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	52	52
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)		
Домашнее задание (ДЗ)	40	40
Контрольные работы заочников (КРЗ)		
Подготовка к промежуточной аттестации (ПА)		
Всего:	92	92

5. Перечень учебно-методического обеспечения

для самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. разделов 6-11.

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

<i>Шифр/ URL адрес</i>	<i>Библиографическая ссылка</i>	<i>Количество экземпляров в библиотеке (кроме электронных экземпляров)</i>
https://e.lanbook.com/book/404567 <i>Режим доступа: для авторизованных пользователей.</i>	Алетдинова, А. А. Интеллектуальный анализ больших данных : учебное пособие / А. А. Алетдинова, М. Ш. Муртазина. — Новосибирск : НГТУ, 2023. — 66 с. — ISBN 978-5-7782-4899-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/404567 (дата обращения: 23.06.2026). — Режим доступа: для авториз. пользователей.	
https://e.lanbook.com/book/427826 <i>Режим доступа: для авторизованных пользователей.</i>	Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие / Т. М. Татарникова. — Вологда : Инфра-Инженерия, 2024. — 172 с. — ISBN 978-5-9729-1772-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/427826 (дата обращения: 23.06.2026). — Режим доступа: для авториз. пользователей.	
https://e.lanbook.com/book/408569 <i>Режим доступа: для авторизованных пользователей.</i>	Ланских, Ю. В. Интеллектуальный анализ данных : учебное пособие / Ю. В. Ланских, В. Г. Ланских. — Киров : ВятГУ, 2023. — 240 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/408569 (дата обращения: 23.06.2026). — Режим доступа: для авториз. пользователей.	

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://pro.guap.ru/	Элементы электронного курса (методические указания к выполнению лабораторных и курсовой работ) по дисциплине размещены внутри ЭИОС ГУАП «Интегрированная среда обучения»

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.	Microsoft Windows 10 (договор ГУАП №1303-3 от 30.12.2019, информация о лицензии представлена по ссылке https://guap.ru/it/system/iso/po)
2.	Microsoft Office 2019 (договор ГУАП №278 от 18.06.2020, информация о лицензии представлена по ссылке https://guap.ru/it/system/iso/po)
3.	Mathcad - (договор ГУАП, информация о лицензии представлена по ссылке https://guap.ru/it/system/iso/po)
4.	MathWorks MATLAB (договор ГУАП, информация о лицензии представлена по ссылке https://guap.ru/it/system/iso/po)
5.	Интегрированная среда программирования Visual Studio Community (свободное использование без ограничений в любой организации для академических исследований; обучения в аудиториях)

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
	Электронные библиотечные ресурсы и системы
1	Электронный каталог библиотеки ГУАП с доступом к базе полнотекстовых изданий (https://lib.guap.ru/), доступ через личный кабинет читателя библиотеки ГУАП
2	Научная электронная библиотека «eLIBRARY» (https://elibrary.ru/), доступ через личный кабинет читателя библиотеки ГУАП, а также по IP -адресам ГУАП
3	ЭБС «Лань» (https://e.lanbook.com/), доступ через личный кабинет читателя библиотеки ГУАП, а также по IP -адресам ГУАП
4	ЭБС Консорциума аэрокосмических вузов России (http://elsau.ru/suai), доступ по IP-адресам ГУАП
5	ЭБС Znanium (https://znanium.ru/), доступ через личный кабинет читателя библиотеки ГУАП, а также по IP -адресам ГУАП
6	образовательная платформа «Юрайт» (https://urait.ru/), доступ через личный кабинет читателя библиотеки ГУАП, а также по IP -адресам ГУАП
7	Научная электронная библиотека «КиберЛенинка» (https://cyberleninka.ru/), свободный доступ
	Информационные и справочно-правовые системы
1	"Консультант Плюс" (www.consultant.ru) сетевая версия для образовательных организаций, доступ по IP -адресам ГУАП

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	Учебная аудитория для занятий лекционного типа. Мультимедийная лекционная аудитория: Специализированная мебель; технические средства обучения, служащие для представления учебной информации большой аудитории; набор демонстрационного оборудования (Интерактивный мультисенсорный дисплей на перекатной стойке FocusTouch Диагональ 70" – 1 шт., ПЭВМ – 1 шт.); Обеспечен доступ в электронную информационно-образовательную среду ГУАП по локальной вычислительной сети	ул.Гастелло, 15
2	Учебная аудитория для практических, лабораторных работ, самостоятельной работы, для групповых и индивидуальных консультаций, для текущего контроля и промежуточной аттестации. Специализированная мебель; лабораторное оборудование: ПЭВМ - 18 шт., объединенных в локальную вычислительную сеть с выходом в вычислительную сеть ГУАП и Интернет; сервер PostgreSQL; экран ScreenMedia GoldView 183*244 MW настенный; сплиттер Kramer VP-200K (с блоком питания)	ул. Гастелло, 15 22-04
	Помещение для самостоятельной работы обучающихся - Читальный зал библиотеки ГУАП: специализированная мебель; персональные компьютеры – 10 шт., обеспечен доступ в электронную информационно-образовательную среду ГУАП по локальной вычислительной сети и точке доступа WiFi, а также к электронно-библиотечным системам, реферативной базе данных Scopus; копировальный аппарат Kyocera KM2035.	ул. Гастелло, 15

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*; Задачи; Тесты.

Примечание: *экзаменационные билеты формируются на основе вопросов и задач таблицы 15.

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 5-балльная шкала	Характеристика сформированных компетенций
«отлично» «зачтено»	Обучающийся: – глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно связывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий. – правильно выполнил от 90% до 100% тестовых заданий**.
«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. – правильно выполнил от 70% до 89% тестовых заданий**.
«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; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий. – правильно выполнил от 51% до 69% тестовых заданий**.
«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. – правильно выполнил менее 51% тестовых заданий**.

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	Большие данные: определение, характеристики (модель «5V») и отличия от традиционных данных.	УК-1.В.2
2	Интеллектуальный анализ данных (Data Mining): место в процессе KDD и основные этапы.	ПК-2.3.1
3	Распределённая файловая система HDFS: архитектура и принципы хранения данных.	УК-1.В.2

4	Модель MapReduce: фазы Map, Shuffle и Reduce для параллельной обработки.	ПК-2.3.1
5	NoSQL базы данных: классификация (документоориентированные, колоночные, графовые, ключ-значение).	УК-1.В.2
6	Data Lake и Data Warehouse: сравнение архитектур и сценариев применения.	ПК-2.3.1
7	CAP-теорема: согласованность, доступность, устойчивость к разделению.	УК-1.В.2
8	Озёра данных (Data Lakes): структура и управление метаданными.	ПК-2.3.1
9	Разведочный анализ данных (EDA): методы визуализации и статистического описания.	УК-1.В.2
10	Обработка пропущенных значений: удаление, импутация, продвинутые методы.	ПК-2.3.1
11	Выбросы и аномалии: методы обнаружения (IQR, Z-score, Isolation Forest).	УК-1.В.2
12	Кодирование категориальных признаков: One-Hot Encoding, Label Encoding, Target Encoding.	ПК-2.3.1
13	Нормализация и стандартизация: Min-Max Scaling, StandardScaler, RobustScaler.	УК-1.В.2
14	Отбор признаков: фильтрационные, обёрточные и встроенные методы.	ПК-2.3.1
15	Метод главных компонент (PCA): снижение размерности и визуализация данных.	УК-1.В.2
16	Балансировка классов: SMOTE, undersampling, class weights.	ПК-2.3.1
17	Деревья решений: критерии разбиения (энтропия, индекс Джини, дисперсия).	УК-1.В.2
18	Случайный лес (Random Forest): принцип бэггинга и случайности признаков.	ПК-2.3.1
19	Градиентный бустинг: XGBoost, LightGBM, CatBoost — особенности и отличия.	УК-1.В.2
20	Метод опорных векторов (SVM): разделяющая гиперплоскость и ядерной трюк.	ПК-2.3.1
21	Логистическая регрессия: сигмоида, функция потерь и интерпретация коэффициентов.	УК-1.В.2
22	Метрики классификации: Accuracy, Precision, Recall, F1-Score, ROC-AUC.	ПК-2.3.1
23	Матрица ошибок (Confusion Matrix): построение и анализ.	УК-1.В.2
24	Метрики регрессии: MSE, RMSE, MAE, R ² и их интерпретация.	ПК-2.3.1
25	Кросс-валидация: K-Fold, Stratified K-Fold, Leave-One-Out.	УК-1.В.2
26	Переобучение и недообучение: диагностика через learning curves и методы борьбы.	ПК-2.3.1
27	Кластеризация K-means: алгоритм, выбор числа кластеров, метод локтя и силуэтный коэффициент.	УК-1.В.2
28	Плотностная кластеризация (DBSCAN): параметры eps и min_samples, обработка шума.	ПК-2.3.1
29	Ассоциативные правила: алгоритмы Apriori и FP-Growth, метрики support, confidence, lift.	УК-1.В.2

30	Рекомендательные системы: контентная и коллаборативная фильтрация, матричные разложения (SVD, ALS).	ПК-2.3.1
----	---	----------

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 выполнения курсового проекта/ курсовой работы представлены в таблице 17.

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

№ п/п	Примерный перечень тем для выполнения курсового проекта/ 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	<p><i>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какая из характеристик НЕ входит в классическую модель «4V» больших данных?</p> <p>a. Volume (объём). b. Velocity (скорость). c. Visualization (визуализация). d. Variety (разнообразие).</p>	ПК-2.3.1
2	<p><i>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какой алгоритм используется для поиска ассоциативных правил на основе свойства антимонотонности?</p> <p>a. K-means. b. Apriori. c. DBSCAN. d. Random Forest.</p>	ПК-2.3.1
3	<p><i>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какая метрика показывает долю верно предсказанных положительных классов от общего числа предсказанных</p>	ПК-2.3.1

	<p>положительных?</p> <p>a. Recall. b. Precision. c. Accuracy. d. F1-Score.</p>	
4	<p><i>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какой тип NoSQL базы данных наиболее эффективен для хранения связей между сущностями (например, социальных сетей)?</p> <p>a. Документоориентированная. b. Колоночная. c. Графовая. d. Ключ-значение.</p>	ПК-2.3.1
5	<p><i>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какой метод снижения размерности основан на сингулярном разложении ковариационной матрицы данных?</p> <p>a. LDA. b. PCA. c. t-SNE. d. UMAP.</p>	ПК-2.3.1
6	<p><i>Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов</p> <p>Какие из перечисленных методов относятся к обучению без учителя?</p> <p>a. K-means. b. Логистическая регрессия. c. DBSCAN. d. Apriori. e. Random Forest.</p>	ПК-2.3.1
7	<p><i>Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора.</i></p> <p>Инструкция: Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов</p> <p>Какие из перечисленных операций являются этапами предобработки данных?</p> <p>a. Вставка пропущенных значений. b. Построение ROC-кривой. c. Нормализация признаков. d. Кодирование категориальных переменных. e. Вычисление F1-Score.</p>	ПК-2.3.1
8	<p><i>Задание комбинированного типа с выбором нескольких вариантов</i></p>	ПК-2.3.1

	<p>ответа из предложенных и развернутым обоснованием выбора. Инструкция: Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов</p> <p>Какие утверждения верны для алгоритма DBSCAN?</p> <p>a. Требуется задания числа кластеров заранее. b. Может обнаруживать кластеры произвольной формы. c. Устойчив к выбросам (шумовые точки помечает отдельно). d. Использует параметры eps и min_samples. e. Основан на итеративном пересчёте центроидов.</p>																	
9	<p>Задание закрытого типа на установление соответствия Инструкция: Прочитайте текст и установите соответствие. К каждой позиции в левом столбце подберите соответствующую позицию в правом столбце.</p> <p>Сопоставьте алгоритмы с типами задач, которые они решают:</p> <table border="0"> <tr> <td>a. K-means;</td> <td>1. Классификация</td> </tr> <tr> <td>b. Apriori;</td> <td>2. Кластеризация.</td> </tr> <tr> <td>c. Random Forest;</td> <td>3. Поиск ассоциативных правил.</td> </tr> <tr> <td>d. Linear Regression;</td> <td>4. Регрессия</td> </tr> </table> <table border="1"> <tr><td>a</td><td></td></tr> <tr><td>b</td><td></td></tr> <tr><td>c</td><td></td></tr> <tr><td>d</td><td></td></tr> </table>	a. K-means;	1. Классификация	b. Apriori;	2. Кластеризация.	c. Random Forest;	3. Поиск ассоциативных правил.	d. Linear Regression;	4. Регрессия	a		b		c		d		ПК-2.3.1
a. K-means;	1. Классификация																	
b. Apriori;	2. Кластеризация.																	
c. Random Forest;	3. Поиск ассоциативных правил.																	
d. Linear Regression;	4. Регрессия																	
a																		
b																		
c																		
d																		
10	<p>Задание закрытого типа на установление соответствия Инструкция: Прочитайте текст и установите соответствие. К каждой позиции в левом столбце подберите соответствующую позицию в правом столбце.</p> <p>Сопоставьте метрики качества классификации с их формулами:</p> <table border="0"> <tr> <td>a. Precision;</td> <td>1. $(TP + TN) / (TP + TN + FP + FN)$</td> </tr> <tr> <td>b. Recall;</td> <td>2. $TP / (TP + FP)$</td> </tr> <tr> <td>c. Accuracy;</td> <td>3. $TP / (TP + FN)$</td> </tr> <tr> <td>d. Specificity;</td> <td>4. $TN / (TN + FP)$</td> </tr> </table> <table border="1"> <tr><td>a</td><td></td></tr> <tr><td>b</td><td></td></tr> <tr><td>c</td><td></td></tr> <tr><td>d</td><td></td></tr> </table>	a. Precision;	1. $(TP + TN) / (TP + TN + FP + FN)$	b. Recall;	2. $TP / (TP + FP)$	c. Accuracy;	3. $TP / (TP + FN)$	d. Specificity;	4. $TN / (TN + FP)$	a		b		c		d		ПК-2.3.1
a. Precision;	1. $(TP + TN) / (TP + TN + FP + FN)$																	
b. Recall;	2. $TP / (TP + FP)$																	
c. Accuracy;	3. $TP / (TP + FN)$																	
d. Specificity;	4. $TN / (TN + FP)$																	
a																		
b																		
c																		
d																		
11	<p>Задание закрытого типа на установление соответствия Инструкция: Прочитайте текст и установите соответствие. К каждой позиции в левом столбце подберите соответствующую позицию в правом столбце.</p> <p>Сопоставьте типы NoSQL баз данных с их примерами:</p> <table border="0"> <tr> <td>a. Документоориентированная;</td> <td>1. Redis</td> </tr> <tr> <td>b. Колоночная;</td> <td>2. MongoDB</td> </tr> <tr> <td>c. Графовая;</td> <td>3. Cassandra</td> </tr> <tr> <td>d. Ключ-значение;</td> <td>4. Neo4j</td> </tr> </table> <table border="1"> <tr><td>a</td><td></td></tr> <tr><td>b</td><td></td></tr> <tr><td>c</td><td></td></tr> <tr><td>d</td><td></td></tr> </table>	a. Документоориентированная;	1. Redis	b. Колоночная;	2. MongoDB	c. Графовая;	3. Cassandra	d. Ключ-значение;	4. Neo4j	a		b		c		d		ПК-2.3.1
a. Документоориентированная;	1. Redis																	
b. Колоночная;	2. MongoDB																	
c. Графовая;	3. Cassandra																	
d. Ключ-значение;	4. Neo4j																	
a																		
b																		
c																		
d																		

12	<p>Задание закрытого типа на установление последовательности. Инструкция: Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо.</p> <p>Установите правильную последовательность этапов процесса KDD (Knowledge Discovery in Databases).</p> <ol style="list-style-type: none"> Интерпретация и визуализация результатов. Выбор и понимание предметной области. Интеллектуальный анализ данных (Data Mining). Предобработка и очистка данных. Постобработка и оценка моделей. <div style="border: 1px solid black; width: 100px; height: 20px; margin-top: 5px;"></div>	ПК-2.3.1
13	<p>Задание закрытого типа на установление последовательности. Инструкция: Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо.</p> <p>Установите последовательность работы алгоритма K-means.</p> <ol style="list-style-type: none"> Пересчёт центроидов кластеров как средних точек. Инициализация центроидов (k случайных точек). Проверка условия останова (сходимость). Назначение каждой точки ближайшему центроиду. <div style="border: 1px solid black; width: 100px; height: 20px; margin-top: 5px;"></div>	ПК-2.3.1
14	<p>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора. Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какая фаза модели MapReduce отвечает за группировку промежуточных данных по ключам и передачу их reducer'ам?</p> <ol style="list-style-type: none"> Map. Shuffle. Reduce. Split. 	ПК-2.3.1
15	<p>Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора. Инструкция: Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p> <p>Какой из перечисленных методов является примером L2-регуляризации и добавляет к функции потерь штраф, пропорциональный сумме квадратов весов модели?</p> <ol style="list-style-type: none"> Lasso-регрессия. Ridge-регрессия (гребневая регрессия). Elastic Net. Dropout. 	ПК-2.3.1
16	<p>Нейрон имеет веса $w=[1,2,-1]$, вход $x=[3,-1,4]$, смещение $b=0.5$. Функция активации — ReLU. Найдите выход нейрона.</p>	УК-1.В.2
17	<p>Нейрон имеет веса $w=[2,-1]$, вход $x=[1,2]$, смещение $b=0$. Функция</p>	УК-1.В.2

	активации — сигмоида. Найдите выход нейрона.	
18	По результатам работы классификатора получена матрица ошибок: TP=80, FP=20, FN=10, TN=90. Найдите точность (Precision) модели.	УК-1.В.2
19	По результатам работы классификатора получена матрица ошибок: TP=45, FP=5, FN=15, TN=35. Найдите полноту (Recall) модели.	УК-1.В.2
20	По результатам классификации 200 объектов: TP=70, FP=10, FN=20, TN=100. Найдите общую точность (Accuracy) модели.	УК-1.В.2
21	Для модели известны: Precision = 0.6, Recall = 0.4. Найдите F1-меру.	УК-1.В.2
22	Признак имеет значения в диапазоне [10;50]. Примените Min-Max нормализацию к значению $x=30$ для приведения к диапазону [0;1].	УК-1.В.2
23	Признак имеет среднее значение $\mu=100$ и стандартное отклонение $\sigma=15$. Найдите стандартизованное значение для $x=130$.	УК-1.В.2
24	Точка данных $x=(3,4)$ и центроид $c=(0,0)$. Найдите евклидово расстояние между ними для определения принадлежности кластеру.	УК-1.В.2
25	В базе транзакций правило «{хлеб} → {масло}» поддерживается 60 транзакциями. При этом «{хлеб}» встречается в 100 транзакциях. Найдите метрику confidence для этого правила.	УК-1.В.2

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
(Ниже приводятся рекомендации по составлению данного раздела)

11.1. Методические указания для обучающихся по освоению лекционного материала (если предусмотрено учебным планом по данной дисциплине).

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:

- Изложение материала по рассматриваемой теме;
- Демонстрация примеров решения конкретных задач;
- Ответы на возникающие вопросы по теме лекции;

11.2. Методические указания для обучающихся по участию в семинарах *(если предусмотрено учебным планом по данной дисциплине)*

Основной целью для обучающегося является систематизация и обобщение знаний по изучаемой теме, разделу, формирование умения работать с дополнительными источниками информации, сопоставлять и сравнивать точки зрения, конспектировать прочитанное, высказывать свою точку зрения и т.п. В соответствии с ведущей дидактической целью содержанием семинарских занятий являются узловые, наиболее трудные для понимания и усвоения темы, разделы дисциплины. Спецификой данной формы занятий является совместная работа преподавателя и обучающегося над решением поставленной проблемы, а поиск верного ответа строится на основе чередования индивидуальной и коллективной деятельности.

При подготовке к семинарскому занятию по теме лекции необходимо ознакомиться с планом его проведения, с литературой и научными публикациями по теме семинара.

11.3. Методические указания для обучающихся по прохождению практических занятий *(если предусмотрено учебным планом по данной дисциплине)*

Практическое занятие является одной из основных форм организации учебного процесса, заключающаяся в выполнении обучающимися под руководством преподавателя комплекса учебных заданий с целью усвоения научно-теоретических основ учебной дисциплины, приобретения умений и навыков, опыта творческой деятельности.

Целью практического занятия для обучающегося является привитие обучающимся умений и навыков практической деятельности по изучаемой дисциплине.

Планируемые результаты при освоении обучающимся практических занятий:

– закрепление, углубление, расширение и детализация знаний при решении конкретных задач;

– развитие познавательных способностей, самостоятельности мышления, творческой активности;

– овладение новыми методами и методиками изучения конкретной учебной дисциплины;

– выработка способности логического осмысления полученных знаний для выполнения заданий;

– обеспечение рационального сочетания коллективной и индивидуальной форм обучения.

Требования к проведению практических занятий изложены в учебно-методическом пособии:

Галанина В.А.Решетов Л.А. Базы данных. Программирование в среде PostgreSQL.- СПб.:ГУАП, 2020 г., 90 с.

11.4. Методические указания для обучающихся по выполнению лабораторных работ *(если предусмотрено учебным планом по данной дисциплине)*

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, Структура и форма отчета о лабораторной работе содержатся в методическом пособии:

004

Г15 В.А.Галанина, Л.А.Решетов. Базы данных. Лабораторный практикум. СПб, ГУАП, 2019 г.

004.6

Г15 Галанина В.А.Решетов Л.А. Базы данных. Программирование в среде PostgreSQL.-СПб.:ГУАП, 2020 г., 90 с.

Задания к лабораторным работам выдаются преподавателем в соответствии с таблицей 6.

11.5. Методические указания для обучающихся по выполнению курсового проекта/ курсовой работы *(если предусмотрено учебным планом по данной дисциплине)*

Курсовой проект/ работа проводится с целью формирования у обучающихся опыта комплексного решения конкретных задач профессиональной деятельности.

Курсовой проект/ работа позволяет обучающемуся:

Не предусмотрено.

11.6. 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.

Для обучающихся по заочной форме обучения выполнение контрольных работ является элементом текущего контроля успеваемости и самостоятельной работы.

В процессе выполнения самостоятельной работы у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет ему 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;
- методические указания по выполнению контрольных работ (для обучающихся по заочной форме обучения).

11.7. 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.

Текущий контроль включает в себя:

- контроль посещаемости;
- защита выполненных лабораторных работ;

В течение семестра обучающиеся загружают в ЭИОС ГУАП отчеты по лабораторным работам в соответствии с установленными НПР требованиями и методами проведения ТКУ, а НПР оценивают загруженные материалы.

Текущий контроль осуществляется в течение семестра в процессе защиты лабораторных работ и выставления соответствующих баллов за работу. Для каждой лабораторной работы указывается предельный срок ее сдачи и максимально возможное количество баллов. В случае сдачи работы после указанного срока, снижается максимально возможное количество баллов за данную работу. В конце семестра подсчитывается общее количество баллов за лабораторные работы и выставляется оценка за лабораторные работы, исходя из следующего правила:

если сумма баллов более 90% от максимальной, оценка «отлично», если 80% -90% - «хорошо», если 60%-80% - «удовлетворительно», если менее 60% - «неудовлетворительно».

11.8. 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».

К экзамену допускаются студенты, сдавшие все лабораторные работы и получившие итоговую положительную оценку.

Оценка за экзамен выставляется по пятибалльной системе в соответствии с таблицей 14.

Аттестационная оценка определяется, как средняя между оценкой за лабораторные работы и оценкой за экзамен. При разнице оценок в один балл, берется оценка за экзамен.

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