

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ЛГПУ»)**

Структурное подразделение Институт физико-математического
образования, информационных и обслуживающих технологий
Кафедра информационных образовательных технологий и систем

УТВЕРЖДАЮ

Директор ИФМОИОТ

Е.Е. Горбенко
«13» декабря 2023 г.



Приложение к рабочей программе учебной дисциплины

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
для проведения текущего контроля и промежуточной аттестации
обучающихся по дисциплине
«Анализ больших данных»**

По направлению подготовки 09.04.04 Программная инженерия
Профиль подготовки Программное обеспечение систем и комплексов
Квалификация выпускника – магистр
Форма обучения очная, заочная
Курс ОФО – 2 курс, ЗФО – 3 курс

Разработчик
Швыров В.В.
канд. физ.-мат. наук, доцент,
доцент кафедры информационных
образовательных технологий и
систем
Заведующий кафедрой
Д.А. Капустин
Протокол от «24» ноября 2023 г. №8

Луганск, 2023

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

1.1. Область применения

Фонд оценочных средств (ФОС) – неотъемлемая часть рабочей программы дисциплины (модуля) Анализ больших данных и предназначен для контроля и оценки образовательных достижений студентов, освоивших программу дисциплины (модуля).

1.2. Цели и задачи фонда оценочных средств

Цель ФОС – установить соответствие уровня подготовки обучающегося требованиям ФГОС ВО бакалавриат / специалитет / магистратура по направлению подготовки 09.04.04 Программная инженерия, утвержденного приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 932 (с изменениями и дополнениями).

1.3. Перечень компетенций, формируемых в процессе освоения основной образовательной программы

Процесс освоения дисциплины направлен на формирование следующих компетенций и индикаторов их достижения:

Код по ФГОС ВО	Индикатор достижения
Универсальные	
Общепрофессиональные	
Профессиональные	
ПК-3. Владение навыками создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов	ПК-3.1. Знать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов ПК-3.2. Уметь использовать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов ПК-3.3. Владеть навыками использования методов создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов

1.4. Этапы формирования компетенций и средства оценивания уровня их сформированности

Этапы формирования компетенций	Компетенции	Контрольно-оценочные средства / способ оценивания
Тема 1. Введение в большие данные.	ПК-3	Выполнение лабораторных работ

Тема 2. Жизненный цикл аналитики данных.	ПК-3	Выполнение лабораторных работ
Тема 3. Высокопроизводительные вычисления.	ПК-3	Выполнение лабораторных работ
Тема 4. Масштабирование и многоуровневое хранение данных.	ПК-3	Выполнение лабораторных работ
Тема 5. Визуализация данных и результатов анализа.	ПК-3	Выполнение лабораторных работ
Тема 6. Классификация задач анализа данных.	ПК-3	Выполнение лабораторных работ
Тема 7. Сложные методы аналитики.	ПК-3	Выполнение лабораторных работ
Тема 8. Основы построения нейронных сетей для анализа данных.	ПК-3	Выполнение лабораторных работ
Текущая аттестация	ПК-3	Контрольная работа
Промежуточная аттестация	ПК-3	Экзамен (письменный)

1.5. Описание показателей формирования компетенций

Код компетенции	Результаты сформированности
ПК-3. Владение навыками создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов	<p>ПК-3.1. Знает методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p> <p>ПК-3.2. Умеет использовать методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p> <p>ПК-3.3. Владеет навыками использования методов создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов</p>

1.6. Критерии оценивания компетенций на разных этапах их формирования

Вид учебной работы	Количество баллов		
3-4 семестр / 7-8 триместр			
	ОФО	О-ЗФО	ЗФО
Оформление отчетов по лабораторным работам	30 баллов		
Работа на лабораторных занятиях	30 баллов		
Выполнение тестовых заданий	-		
Выполнение заданий самостоятельной	10 баллов		

работы			
экзамена	30 баллов		
Итого за семестр:	100 баллов		
Всего	100 баллов		

Накопительная система оценивания по 100-балльной шкале

Четырехбал- льная система оценивания экзамена	100- балльная шкала	Буквенная шкала, соответствующая 100- балльной шкале	Система оценивания зачета
Отлично	90–100	А – отлично – теоретическое содержание курса освоено полностью, без пробелов; необходимые практические навыки работы с освоенным материалом сформированы; все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному	Зачтено
Хорошо	83–89	В – очень хорошо – теоретическое содержание курса освоено полностью, без пробелов; необходимые практические навыки работы с освоенным материалом в основном сформированы; все предусмотренные программой обучения учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному	
Хорошо	75–82	С – хорошо – теоретическое содержание курса освоено полностью; некоторые практические навыки работы с освоенным материалом сформированы недостаточно; все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками	
Удовлетво- рительно	63–74	Д – удовлетворительно – теоретическое содержание дисциплины освоено частично, но пробелы не носят существенного характера; необходимые практические навыки работы с освоенным материалом в основном сформированы; большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, содержат ошибки	
Удовлетво- рительно	50–62	Е – посредственно – теоретическое содержание курса освоено частично; некоторые практические навыки работы не сформированы, многие предусмотренные программой обучения учебные задания не выполнены либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному	

Неудовлетворительно	21–49	FX – неудовлетворительно – теоретическое содержание курса освоено частично; необходимые практические навыки работы не сформированы; большинство предусмотренных программой обучения учебных заданий не выполнено либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий	Не зачтено
Неудовлетворительно	0–20	F – неудовлетворительно – теоретическое содержание курса не освоено; необходимые практические навыки работы не сформированы; все выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над материалом курса не приведет к какому-либо значимому повышению качества выполнения учебных заданий	

2. КОНТРОЛЬНО-ОЦЕНОЧНЫЕ СРЕДСТВА

2.1. Оценочные средства текущего контроля (типовые)

Вопросы для текущего контроля:

1. Что представляют собой большие данные (Big Data)?
2. Какие основные характеристики определяют данные как "большие"?
3. Какие три основных аспекта характеризуют большие данные по модели "три В"?
4. Какие вызовы и проблемы возникают при работе с объемными данными?
5. Какие типы данных обычно считаются частью больших данных?
6. Каким образом большие данные отличаются от традиционных баз данных?
7. Как технологии обработки больших данных влияют на аналитику и принятие решений?
8. Какие технологии хранения данных часто используются для больших данных?
9. Каким образом обработка больших данных может влиять на производительность вычислительных систем?
10. Какие преимущества предоставляет облачные вычисления для обработки больших данных?
11. Какие технологии распределенных вычислений часто применяются для работы с большими данными?
12. Как обеспечивается безопасность при обработке и хранении больших данных?
13. Каким образом большие данные влияют на сферу бизнеса и промышленности?
14. Какие методы анализа больших данных используются для выявления закономерностей и трендов?
15. Как происходит процесс сбора и загрузки больших данных в хранилища?
16. Каким образом большие данные могут быть использованы в машинном обучении и искусственном интеллекте?
17. Какие вызовы возникают при обработке и анализе структурированных и неструктурированных данных?
18. Как обработка потоковых данных связана с темой больших данных?
19. Какие методы визуализации данных применяются для больших данных?
20. Какие роли и навыки требуются у специалистов, работающих с большими данными?
21. Каким образом концепция "Data Lake" связана с хранением и обработкой больших данных?
22. Какие вызовы возникают при обеспечении качества данных в больших объемах?
23. Каким образом технологии распределенного хранения данных решают проблемы масштабирования?

24. Какие примеры успешного использования больших данных существуют в различных отраслях?
25. Какие тенденции и будущие направления развития области больших данных можно выделить?
26. Что означает термин "масштабирование данных" в контексте информационных систем?
27. Какие преимущества масштабирования данных на уровне хранения обеспечиваются при увеличении объема информации?
28. Как различаются вертикальное и горизонтальное масштабирование данных?
29. Какие вызовы существуют при вертикальном масштабировании баз данных?
30. Что представляет собой горизонтальное масштабирование баз данных, и как оно может быть реализовано?
31. Какие архитектурные подходы к масштабированию данных применяются в распределенных системах?
32. Какие роли играют шардирование и репликация при горизонтальном масштабировании данных?
33. Как технологии кэширования могут быть использованы для улучшения производительности распределенных систем?
34. Как масштабирование данных связано с обработкой потоковых данных и реального времени?
35. Какие тенденции существуют в области технологий масштабирования баз данных?
36. Какие решения для масштабирования NoSQL баз данных широко используются в индустрии?
37. Каким образом технологии кластеризации и контейнеризации влияют на масштабирование данных?
38. Какие вызовы возникают при обеспечении целостности данных в распределенных системах масштабирования?
39. Какие методы балансировки нагрузки применяются при горизонтальном масштабировании баз данных?
40. Какие технологии обеспечивают высокую доступность и отказоустойчивость при масштабировании данных?
41. Каким образом многоуровневое хранение данных влияет на эффективность и оптимизацию использования ресурсов?
42. Какие технологии хранения данных на разных уровнях (hot, warm, cold storage) позволяют экономить ресурсы?
43. Какие решения для архивирования данных применяются в контексте многоуровневого хранения?
44. Как автоматизация управления данными влияет на процессы масштабирования и хранения данных?
45. Какие вызовы и тенденции существуют в области секьюрити при масштабировании и многоуровневом хранении данных?
46. Каким образом аналитические базы данных поддерживают масштабирование для обработки больших объемов данных?

47. Какие методы агрегации и компрессии данных применяются при их перемещении между уровнями хранения?
48. Как технологии управления жизненным циклом данных согласуются с масштабированием и многоуровневым хранением?
49. Как масштабирование и многоуровневое хранение данных согласуются с требованиями по соблюдению законодательства о данных?
50. Какую роль играют решения в области обработки больших данных (Big Data) в стратегиях масштабирования и хранения данных?
51. Какие библиотеки в Python чаще всего используются для работы с большими данными?
52. Каким образом библиотека Pandas облегчает работу с большими данными в Python?
53. Как использование библиотеки NumPy может улучшить производительность операций над большими массивами данных в Python?
54. Какие методы в библиотеке Matplotlib применяются для визуализации больших данных в Python?
55. Как библиотека Dask помогает в обработке и анализе данных, не помещающихся в оперативной памяти?
56. Как распределенные вычисления с использованием библиотеки Apache Spark могут быть реализованы в Python?
57. Каким образом библиотека Vaex позволяет эффективно работать с большими наборами данных?
58. Какие методы в библиотеке Scikit-learn используются для машинного обучения на больших данных в Python?
59. Как библиотека TensorFlow обеспечивает возможности работы с большими объемами данных для глубокого обучения?
60. Какие инструменты Python могут использоваться для обработки потоковых данных в режиме реального времени?
61. Как библиотека Apache Flink интегрируется с Python для обработки потоков данных?
62. Каким образом библиотека GeoPandas поддерживает обработку и анализ географических данных в Python?
63. Как использование библиотеки Plotly может улучшить интерактивную визуализацию больших данных в Python?
64. Как библиотека Arrow помогает в эффективной обработке временных рядов и дат в больших данных?
65. Каким образом библиотека Joblib облегчает распределенные вычисления и параллельную обработку данных в Python?
66. Как библиотека Psycopg2 поддерживает взаимодействие с базами данных PostgreSQL для работы с большими объемами данных?
67. Каким образом библиотека PySpark интегрируется с языком программирования Python для анализа данных на кластерах?
68. Какие методы оптимизации работы с памятью применяются при работе с большими данными в Python?
69. Как библиотека Hadoop MapReduce может использоваться в связке с Python для обработки больших данных?

70. Каким образом библиотека Scrapy поддерживает сбор и обработку данных с веб-сайтов в больших масштабах?
71. Как библиотека SQLAlchemy обеспечивает доступ и взаимодействие с реляционными базами данных в Python?
72. Каким образом библиотека Faust поддерживает обработку потоковых данных с использованием асинхронного программирования в Python?
73. Какие вызовы могут возникнуть при работе с распределенными системами хранения данных в Python?
74. Как использование библиотеки PyTorch может быть ценным при обработке и анализе больших данных в области искусственного интеллекта?
75. Какие стратегии кэширования и предварительной обработки данных используются при работе с большими объемами информации в Python?

2.2. Оценочные средства для промежуточной аттестации

Вопросы для проведения аттестации

1. Какие цели преследует визуализация данных в контексте анализа?
2. Как использование различных типов графиков может поддерживать разные аспекты визуализации данных?
3. Какие библиотеки Python часто применяются для создания визуализаций?
4. Какие основные элементы включаются в график для передачи информации?
5. Как использование цвета и стиля помогает в передаче информации в визуализации данных?
6. Каким образом визуализация временных рядов может помочь в анализе данных?
7. Какие вызовы могут возникнуть при визуализации больших объемов данных?
8. Какие типы диаграмм подходят для отображения распределения данных?
9. Как использование тепловых карт может помочь в анализе плотности данных?
10. Как визуализация диаграммы рассеяния может помочь в выявлении корреляции между переменными?
11. Каким образом визуализация облака точек используется для представления многомерных данных?
12. Какие методы визуализации данных подходят для категориальных переменных?
13. Каким образом создание диаграммы "ящик с усами" (box plot) может помочь в выявлении статистических характеристик данных?
14. Какие инструменты и библиотеки обеспечивают интерактивную визуализацию данных?
15. Каким образом графики Ганта используются для отображения временных данных и событий?

16. Как визуализация сетевых графиков может помочь в анализе взаимосвязей между элементами?
17. Как использование дашбордов и интерактивных отчетов облегчает представление результатов анализа?
18. Каким образом тематическая картография поддерживает визуализацию пространственных данных?
19. Как визуализация данных может быть адаптирована для лиц, не обладающих навыками анализа?
20. Как использование анимации в визуализации данных может улучшить восприятие временных изменений?
21. Как визуализация данных может помочь в обнаружении выбросов и аномалий?
22. Каким образом визуализация данных может поддерживать процессы машинного обучения, такие как кластеризация?
23. Какие вызовы существуют при визуализации данных с использованием виртуальной и дополненной реальности?
24. Каким образом выбор подходящих цветовых схем влияет на читаемость и эффективность визуализации?
25. Как использование динамической визуализации данных может помочь в анализе изменяющихся паттернов в данных?
26. Что представляют собой сложные методы аналитики данных и в чем их отличие от базовых методов?
27. Каким образом методы глубокого обучения применяются для анализа и обработки сложных структурированных данных?
28. Какие вызовы существуют при работе с неструктурированными данными, и какие методы аналитики могут быть эффективны в этом случае?
29. Как технологии обработки естественного языка (NLP) применяются для анализа текстовых данных?
30. Как методы временных рядов и анализа временных последовательностей используются для прогнозирования и выявления паттернов в данных?
31. Каким образом методы кластеризации применяются для выявления групп схожих объектов в данных?
32. Какие методы анализа ассоциаций применяются для выявления связей между различными переменными в данных?
33. Каким образом алгоритмы обучения без учителя, такие как t-SNE, используются для визуализации сложных многомерных данных?
34. Какие методы анализа данных применяются для обнаружения выбросов и аномалий в данных?
35. Как алгоритмы рекомендательных систем могут быть использованы для предсказания предпочтений пользователей?
36. Какие методы временных рядов и анализа временных последовательностей используются для прогнозирования и выявления паттернов в данных?
37. Каким образом методы машинного обучения, такие как ансамблевые методы, применяются для повышения точности анализа данных?

38. Каким образом байесовские методы применяются в аналитике данных для моделирования вероятностей событий?
39. Какие методы анализа данных применяются для извлечения признаков и уменьшения размерности данных?
40. Как алгоритмы обучения с подкреплением применяются для принятия решений в сложных средах?
41. Каким образом методы геоинформационного анализа применяются для работы с пространственными данными?
42. Как методы генетического программирования используются для оптимизации сложных моделей аналитики данных?
43. Каким образом алгоритмы обучения на основе примеров (exemplar-based learning) применяются для анализа данных?
44. Какие методы анализа данных используются для работы с изображениями и обработки видео?
45. Каким образом методы временных рядов применяются для анализа сезонности и цикличности в данных?
46. Как алгоритмы обучения с подкреплением применяются для принятия решений в сложных средах?
47. Каким образом методы генетического программирования используются для оптимизации сложных моделей аналитики данных?
48. Как методы анализа сетей и графов применяются для выявления взаимосвязей в сложных системах?
49. Как алгоритмы обучения на основе примеров (exemplar-based learning) применяются для анализа данных?
50. Каким образом методы временных рядов применяются для анализа сезонности и цикличности в данных?
51. Что представляют собой искусственные нейронные сети?
52. Каким образом нейроны в нейронной сети моделируют биологические нейроны?
53. Какие типы слоев чаще всего встречаются в нейронных сетях?
54. Как функция активации влияет на выход нейрона в нейронной сети?
55. Каким образом определяются веса в нейронной сети и как они обучаются?
56. Как происходит процесс прямого распространения (forward propagation) в нейронной сети?
57. Что представляют собой функции потерь (loss functions) в контексте обучения нейронных сетей?
58. Как работает обратное распространение ошибки (backpropagation) в процессе обучения нейронной сети?
59. Какие регуляризационные методы могут применяться для предотвращения переобучения в нейронных сетях?
60. Что такое оптимизаторы в контексте обучения нейронных сетей и какие они бывают?
61. Как выбор функции активации может зависеть от задачи, которую решает нейронная сеть?
62. Как определяется архитектура нейронной сети, включая количество слоев и количество нейронов в каждом слое?

63. Какие гиперпараметры влияют на обучение нейронной сети, и как их подбирать?
64. Какие методы и техники инициализации весов применяются при создании нейронных сетей?
65. Каким образом нейронные сети используются для решения задачи регрессии?
66. Какие возможности предоставляет нейронным сетям решение задачи классификации?
67. Что такое сверточные нейронные сети (CNN) и в каких областях они часто применяются?
68. Как работают рекуррентные нейронные сети (RNN) и в каких задачах они эффективны?
69. Каким образом архитектура сети влияет на возможность распознавания сложных паттернов в данных?
70. Какие методы передачи обучения (transfer learning) могут использоваться при построении нейронных сетей?
71. Каким образом решаются проблемы нестационарности данных в нейронных сетях?
72. Какие вызовы возникают при интерпретации результатов нейронных сетей?
73. Каким образом определить оптимальные гиперпараметры для нейронной сети?
74. Как работают автокодировщики и для чего они могут использоваться в анализе данных?
75. Каким образом решаются проблемы нестационарности данных в нейронных сетях?