Министерство науки и высшего образования
Российской Федерации
Министерство науки, высшего образования и инноваций

Межгосударственная образовательная организация высшего образования Кыргызско - Российский Славянский университет имени первого Президента Российской Федерации Б.Н. Ельцина

Кафедра Информационных и вычислительных технологий

Фонд
оценочных средств

по дисциплине «Аналитика больших данных»

Уровень высшего образования

МАГИСТРАТУРА

Направление подготовки

09.04.04 - РФ, 710400 - КР Программная инженерия

(код и наименование направления подготовки)

Разработка программно-информационных систем

(профиля) образовательной программы)

Квалификация

магистр

2025

Фонд оценочных средств предназначен для контроля знаний обучающихся по направлению подготовки 09.04.04 – РФ, 710400 – КР «Программная инженерия» по дисциплине
«Аналитика больших данных».

Фонд оценочных средств рассмотрен и утвержден на заседании Информационных и вычислительных технологий протокол № 1 от 03.09.2025 г.

Заведующий кафедрой
д.т.н., проф. Лыченко Н.М.

Исполнители (разработчики):
К.т.н., доцент каф. ИВТ Верзунов С.Н.
К.т.н., доцент каф. ИВТ Манжикова С.Ц.

СОГЛАСОВАНО:

И.о. декана ЕТФ Н.М. Комарцов

Раздел 1. Перечень компетенций, с указанием этапов их формирования в процессе освоения дисциплины/практики

Формируемые компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Виды оценочных средств/ шифр раздела в данном документе
ПК-3: Способен выполнить постановку задач анализа и синтеза проектных решений	**Знать:** основные технологии построения программных систем: процедурные подходы, объектные и компонентные модели, базовые шаблоны проектирования программных систем.	***Блок А –*** *задания репродуктивного уровня* - *Тест*
	Уметь: - применять адекватные методы и средства для реализации прикладных задач и разработки реальных программных систем с целью анализа и синтеза проектных решений.	**Блок В** – задания реконструктивного уровня - Практические задания
	Владеть: практическими навыками работы с современными системами проектирования и разработки программных систем.	**Блок С** – задания практико ориентированного и/или исследовательского уровня - Практические задания
ПК-4: Спосо применять современные технологии разработки программных комплексов, осуществлять контроль качества разрабаты-ваемых программных продуктов	**Знать:** методы и средства создания программного обеспечения на основе современных компьютерных технологий.	**Блок А –** задания репродуктивного уровня - Тест
	Уметь: использовать современные компьютерные технологии в разработке программных комплексов.	**Блок В** – задания реконструктивного уровня - Практические задания
	Владеть: современными технологиями разработки программных комплексов.	**Блок С -** задания практико ориентированного и/или исследовательского уровня - Практические задания

Раздел 2. Технологическая карта дисциплины Аналитика Больших Данных

Курс 1, семестр 2, Количество ЗЕ -3, Отчетность – зачет

Название модулей дисциплины согласно РПД	Контр	Форма контро	зачетный минимум	зачетный максимум	График кон
1. Введение в большие данные	текущий	Сдача практических №1	5	8	26
	рубежный	Тест №1	5	8	26
2. Жизненный цикл проекта по аналитике больших данных	текущий	Сдача практических №2	5	9	29
	рубежный	Тест №2	5	9	29
3. Основные техники работы с большими данными	текущий	Сдача практических №3, №4, №5	5	9	33
	рубежный	Тест №3	5	9	33
4. Приложение больших данных	текущий	Сдача практических №6	5	9	35
	рубежный	Тест №4	5	9	35
ВСЕГО за семестр			40	70	
Промежуточный контроль (Экзамен)			20	30	
Семестровый рейтинг по дисциплине			60	100	

Раздел 3. Типовые контрольные задания и иные материалы, необходимые для оценки планируемых результатов обучения по дисциплине (оценочные средства)

Блок А

1. Большинство данных в мире в 2011 году содержалось:
   i. В цифровом виде
   ii. В аналоговом виде

2. Объём накопленных человечеством цифровых данных на 2012 год измеряется:
   i. Петабайтами
   ii. Зеттабайтами
   iii. Экзабайтами
   iv. Йоттабайтами

3. Сколько Петабайт в Зеттабайте?

4. Укажите фактор, способствовавший появлению тренда больших данных
   i. Маркетинговые кампании крупных корпораций

ii. Снижение издержек на хранение данных
iii. Появление новых технологий обработки потоковых данных
iv. Выпуск баз данных с обработкой данных в памяти

5. Какие вероятные разочарования тренда больших данных?
i. Из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных.

6. Отметьте значимые события, повлиявшие на формирование тренда больших данных:
i. Разработка Hadoop
ii. Изобретение принципа MapReduce
iii. Разработка языка Pyhton
iv. Победа Deepblue в матче с Г.Каспаровым.

7. Определение больших данных:
i. Большие данные – это обработка или хранение более 1 Тб информации.
ii. Проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки сущност-ная обработка данных затруднена или невозможна.
iii. Большие данные – это огромная PR-акция крупных вендоров и не более того.
iv. Большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект.

8. Выберите неверный ответ:
v. Большие данные – это данные объёма свыше 1 Тб
vi. Проблема больших данных – это проблема, когда при суще-ствующих технологиях хранения и обработки сущностная об-работка данных затруднена или невозможна.
vii. Большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров.
viii. Большие данные как правило не структурированы.

9. Отметьте те из вариантов, в которых данные структурированы:
ix. Данные о продажах компании, представленные в виде поме-сячных отчётов в формате MS Word.
x. Таблица с ежедневными показаниями температуры помеще-ния за год в файле формата csv.
xi. Текст педагогической поэмы А.С. Макаренко, представлен-ный в формате PDF.
xii. Библиотека фильмов, представленных в формате mpeg4 на одном жестком диске.

10. Перечислите четыре основных характеристики Big Data:
xiii. Virtualization, Volume, Variability, Vehicle

xiv. Variety, Velocity, Volume, Value
xv. Verification, Volume, Velocity, Visualization
xvi. Video, Value, Variety, Volume

11Выберите неверное высказывание:
i. Большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных.
ii. Увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации.
iii. Удешевление систем хранения на единицу информации при-вело к росту рынка больших данных.
iv. Большое разнообразие источников данных

12. Отметьте неверное понимание Variety в контексте характеристик Big Data:
i. Высокая скорость генерирования данных.
ii. Разные типы данных в колонках таблиц реляционных СУБД.
iii. Разнообразие отраслей, являющихся источниками данных.
iv. Разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные.

13. Принцип MapReduce состоит в том, чтобы
i. Производить вычисления на узлах, где информация изначально была сохранена
ii. Использовать вычислительные мощности систем хранения
iii. Использовать функциональное программирование для решения задач массивно-параллельной обработки

14. Выберите одно неверное высказывание про MapReduce:
iv. Интерфейс для массово-параллельной обработки данных, где вычисления производятся на узлах, где информация изна-чально была сохранена
v. MapReduce – это две операции: распределения и сборки данных
vi. MapReduce был придуман разработчиками Hadoop
vii. MapReduce был анонсирован разработчиками Google

15. Каков теоретический прирост производительности при подсчёте числа слов в тексте при работе MapReduce при переходе от одного узла к двум?

16. Какие из следующих технологий СУБД не используют принцип MapReduce
viii. Hadoop
ix. Cassandra
x. HDInsight
xi. Redis

17. Какие СУБД полностью полагаются на оперативную память при хранении информации:

xii. Oracle Exalytics
xiii. SAP HANA
xiv. BigTable
xv. HBase

18. В чём преимущество колоночно-ориентированных СУБД?

i. Они позволяют выполнять более сложные SQL-запросы по сравнению с реляционными СУБД
ii. Они позволяют динамически дополнять содержание записей новыми полями
iii. Они имеют более гибкие возможности аналитики.
iv. Они позволяют эффективно делать межколоночные сравнения.

19. Для чего аналитику необходима «песочница»?

i. Для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций.
ii. Для хранения всех полученных от заказчика данных.
iii. Для построения отчётов о результатах анализа
iv. Для снижения затрат, связанных с репликацией данных

20. Какие из следующих средств разумно использовать для анализа данных, представленных единственным csv-файлом размера более 100Гб:

v. Hadoop
vi. Data Warehouse
vii. «Песочница»
viii. Python

21. Выберите верное утверждение:

ix. Data Warehouse создаются для проверки гипотез при анализе больших данных.
x. «Песочница» используется для снижения нагрузки на основной Data Warehouse.
xi. Каждый Data Warehouse должен содержать «песочницу».
xii. «Песочница» необходима для любого процесса аналитики.
xiii.

21. Расставьте последовательность этапов проекта аналитики в соответствии с CRISP-DM.

xiv. Понимание бизнеса (Business understanding)
xv. Понимание данных (Data Understanding)
xvi. Подготовка данных (Data Preparation)
xvii. Моделирование (Modeling)
xviii. Оценка (Evaluation)
xix. Внедрение (Deployment)

22. На каком из этапов процесса CRISP-DM происходит проверка гипотез?

xx. Понимание бизнеса (Business understanding)
xxi. Понимание данных (Data Understanding)
xxii. Моделирование (Modeling)
xxiii. Оценка (Evaluation)

23. Вы являетесь владельцем и аналитиком в компании из 10 человек, в которой требуется проанализировать продажи за 1 год (1 млн. продаж). Какие из этапов CRISP-DM можно опустить:

i. Понимание бизнеса (Business understanding)
ii. Подготовка данных (Data Preparation)
iii. Моделирование (Modeling)
iv. Оценка (Evaluation)

24. Пример благоразумного использования Hadoop

i. Анализ 10 Гб данных.
ii. Ежедневное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт).
iii.Посекундное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт).
iv. Построение графика пульса пациента в реальном времени.

25. Начиная с каких размеров данных обоснованно применение кластера Hadoop для хранения данных?

v. 100Гб
vi. 1Тб
vii. 100Тб
viii. 1Пб

26. Hadoop – это:

ix. Набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах.
x. Распределённая СУБД, позволяющая обрабатывать большие данные.
xi. Язык выполнения заданий в парадигме MapReduce.
xii. Распределённая файловая система, предназначенная для хранения файлов большого объёма.

Блоки B и C

Практические задания

Практическая работа 1. Библиотека NumPy

Задания:

1. Загрузите массив из файла
2. Измените форму массива
3. Упорядочите элементы массива

4. Разбейте и объедините массив

Практическая работа 2. Анализ набора данных usa.gov data from Bitly

2011 году служба сокращения URL-адресов bit.ly заключила партнерское соглашение с сайтом правительства США USA.gov (https://www.usa.gov/) о син-хронном предоставлении анонимных данных о пользователях, которые со-кращают ссылки, заканчивающиеся на .gov или .mil. В 2011 году, помимо синхронной ленты, формировались ежечасные мгновенные снимки, доступ-ные в виде текстовых файлов. В 2017 году эта служба уже закрылась, но мы сохранили файлы данных и приводим их в качестве примеров.

Задания

1. Загрузите данные из файла
2. Подсчитайте количество часовых поясов

Практическая работа 3. Анализ набора данных MovieLens 1M

Исследовательская группа GroupLens Research (https://grouplens.org/datasets/ movielens/) предлагает несколько наборов данных о рейтингах фильмов, проставленных пользователями сайта MovieLens в конце 1990-х – начале 2000-х годов. Наборы содержат рейтинги фильмов, метаданные о фильмах (жанр и год выхода) и демографические данные о пользователях (возраст, почтовый индекс, пол и род занятий). Такие данные часто представляют интерес для разработки систем рекомендования, основанных на алгоритмах машинного обучения. И хотя в этой книге методы машинного обучения не рассматриваются, я все же покажу, как формировать продольные и попереч-ные разрезы таких наборов данных с целью привести их к нужному виду.

Задания

1. Загрузите данные из zip-файла
2. Выведите начало и конец данных
3. Выполните статистический анализ данных
4. Измерьте несогласия в оценках

Практическая работа 4. Анализ имен, которые давали детям в США за период с 1880 по 2010 год

Управление социального обеспечения США выложило в Сеть данные о часто-те встречаемости детских имен за период с 1880 года по настоящее время. Хэдли Уикхэм (Hadley Wickham), автор нескольких популярных пакетов для R, часто использует этот пример для иллюстрации манипуляций с данными в R.

Задания

1. Создайте набор данных
2. Постройте график общего количество родившихся в зависимости от пола и года рождения

3. Проанализируйте тенденции в выборе имен

Практическая работа 5. База данных о продуктах питания министерства сельского хозяйства США

Министерство сельского хозяйства США публикует данные о пищевой цен-ности продуктов питания. Программист Эшли Уильямс (Ashley Williams) пре-образовал эту базу данных в формат JSON. Записи выглядят следующим образом

```
  {"id": 21441,
  "description": "KENTUCKY FRIED CHICKEN, Fried Chicken, EXTRA CRISPY, Wing, meat and skin with breading",
  "tags": ["KFC"],
  "manufacturer": "Kentucky Fried Chicken",
  "group": "Fast Foods",
  "portions": [
    {
      "amount": 1,
      "unit": "wing, with skin",
      "grams": 68.0
    },
    ...
  ],
  "nutrients": [
    {
      "value": 20.8,
      "units": "g",
      "description": "Protein",
      "group": "Composition"
    },
    ...
  ]
}
```

Для каждого продукта питания есть ряд идентифицирующих атрибутов и два списка: питательные элементы и размеры порций. Для анализа данные в та-кой форме подходят плохо, поэтому необходимо их переформатировать.

Задания

1. Загрузите архив с сайта
2. Покажите распределение продуктов питания по группам
3. Постройте медианные значения цинка по группе питательных элементов

Практическая работа 6. Анализ базы данных федеральной избирательной комиссии

Федеральная избирательная комиссия США публикует данные о пожертво-ваниях участникам политических кампаний. Указывается имя жертвователя, род занятий, место работы и сумма пожертвования. Интерес представляет на-бор данных, относящийся к президентским выборам 2012 года.

Задания

1. Загрузите набор данных из файла
2. Покажите статистику пожертвований по роду занятий и месту работы
3. Постройте график Общая сумма пожертвований по партиям для родов занятий с максимальной суммой пожертвований
4. Выполните статистический анализ данных
5. Измерьте несогласия в оценках

Блок D (промежуточный контроль)

Вопросы для проверки уровня обученности ЗНАТЬ

1. Основные характеристики больших данных.
2. Роль больших данных в экономике.
3. Основные конструкции языка R.
4. HDFS - основы организации.
5. Архитектура Hadoop.
6. Исполнение Map/Reduce.Исполнение программ в Hadoop.Основы YARN.
7. Аналитика потоковых данных в платформе Storm.
8. Архитектура Apache Spark.Алгоритмы классификации.
9. Алгоритмы кластеризации.
10. Нейронные сети как реализация алгоритмов машинного обучения.
11. Предиктивные алгоритмы.
12. Применение технологий больших данных для задач управления в реальном времени.
13. Понятие бизнес-аналитики.
14. Классификация средств «бизнес-аналитики».
15. Этапы анализа данных. KDD.
16. Общая характеристика QlikView.
17. Общая характеристика DeductorAcademic.
18. Общая характеристика средств интеллектуального анализа SQLServer.
19. Консолидация данных.
20. Визуализация данных.
21. Системы поддержки принятия решений.
22. Хранилища данных.
23. Размерностные модели.
24. OLAP-куб.
25. Таблица размерностей.
26. Таблица фактов.
27. Сравнительный анализ OLAP и OLTP-систем.
28. Консолидация данных в R. Визуализация в R.
29. Организация данных в Apache Spark.
30. Обработка данных в GraphX.Data Mining. Сдерства обработки Data Mining.
31. Элементы математической статистики.

32. Описательная статистика.
33. Операции агрегирования данных.
34. Графические средства анализа.
35. Диаграмма рассеяния.
36. Гистограмма.
37. Начальные этапы KDD. ETL.
38. Средства очистки и трансформации данных.
39. Классификация - метод предобработки и очистки данных.
40. Методы борьбы с аномалиями.
41. Ящечная диаграмма.
42. Общая характеристика задач кластерного анализа.
43. Метрики кластерного анализа.
44. Методы определения близости между кластерами.
45. Иерархическая кластеризация.
46. Дендограмма.
47. Метод к-средних.
48. Определение регрессионной модели.
49. Логистическая регрессионная модель.
50. Ассоциативные правила.
51. Свойства антимонотонности.
52. Метрики построения ассоциативных правил.

Задания для проверки уровня обученности УМЕТЬ и ВЛАДЕТЬ

1. Постройте график Общая сумма пожертвований по партиям для родов занятий с максимальной суммой пожертвований
2. Покажите распределение продуктов питания по группам
3. Постройте медианные значения цинка по группе питательных элементов
4. Постройте метрики Ассоциативных правил
5. Определите близость между кластерами.
6. Постройте регрессионную модель
7. Постройте диаграмму рассеяния и гистограмму
8. Постройте OLAP-куб

Пример

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ ПРОМЕЖУТОЧНОГО КОНТРОЛЯ № ___

1. Вопросы для проверки уровня обученности ЗНАТЬ
   1.1 Основные конструкции языка R.
   1.2 Общая характеристика DeductorAcademic.

2. Задание для проверки уровня обученности УМЕТЬ и ВЛАДЕТЬ
   Определите близость между кластерами.

Раздел 4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Применяемые оценочные средства:

- Сдача практических работ на практических занятиях в соответствии с технологической картой дисциплины (текущая и рубежная аттестация),
- Тест;
- Письменный опрос по экзаменационным билетам (промежуточная аттестация - зачет с оценкой).

Все виды оценочных средств оцениваются в соответствии со шкалами оценивания.

Устный опрос на практических занятиях по отдельным темам проводится в течение всего периода обучения дисциплине. Результаты опроса учитываются при оценивании практических работ.

ШКАЛА ОЦЕНИВАНИЯ ТЕСТА (рубежный контроль)

- В одном тестовом задании некоторое количество закрытых вопросов.
- К заданиям даются готовые ответы на выбор, один правильный и остальные неправильные.
- Обучающемуся необходимо помнить: в каждом задании с выбором одного правильного ответа правильный ответ должен быть.
- За каждый правильно ответ – 5 баллов
- Определяется сумма набранных баллов.
- Вычисляется отношение набранных баллов к максимально возможным за тест. Отметка выражается в %.

ШКАЛА ОЦЕНИВАНИЯ ПРАКТИЧЕСКИХ (ЛАБОРАТОРНЫХ) РАБОТ
(текущий контроль)

- 85-100 % - Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию, выполнены.
- 70-84 % - Демонстрирует значительное понимание проблемы. Все требования, предъявляемые к заданию, выполнены.
- 60-69 % - Демонстрирует частичное понимание проблемы. Большинство требований, предъявляемых к заданию выполнены.
- 31-60 % - Демонстрирует небольшое понимание проблемы. Многие требования, предъявляемые к заданию не выполнены.
- 0-30 % - Демонстрирует непонимание проблемы и даже не было попытки решить задачу.

ШКАЛА ОЦЕНИВАНИЯ ПИСЬМЕННОГО ОПРОСА

(**промежуточный контроль** – «ЗНАТЬ»)

Отметкой (7-10- баллов) оценивается ответ, который показывает прочные знания теоретических основ дисциплины, понимание и правильное применение терминологии, правильные ответы на 75-100% вопросов

Отметкой (5-7 баллов) оценивается ответ, который показывает знание теоретических основ дисциплины, но неполное понимание и не всегда правильное применение терминологии, даны правильные ответы на 50-74% вопросов, в ответах допущено некоторое количество неточностей.

Отметкой (3-4 баллов) оценивается ответ, свидетельствующий о знакомстве с некоторыми теоретическими основами дисциплины. Даны правильные ответы на 25-49% вопросов, допущены неточности и ошибки.

Отметкой (2 балла) оценивается ответ, обнаруживающий незнание теоретических основ дисциплины. Отмечается отсутствие логичности и последовательности в ответе. Менее 25% правильных ответов. Допущены серьезные ошибки в содержании ответа.
Отметкой (0-1 балл) оценивается ответ, при котором студент демонстрирует непонимание поставленных вопросов, или нет ответа.

ШКАЛА ОЦЕНИВАНИЯ ПРАКТИЧЕСКИХ ЗАДАНИЙ
(**промежуточный контроль** – «УМЕТЬ и ВЛАДЕТЬ»)
Отметкой (8-10 баллов) оценивается ответ, при котором студент правильно отвечает на поставленные вопросы, Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию, выполнены.
Отметкой (5-7 баллов) оценивается ответ, при котором студент в основном правильно отвечает на поставленные вопросы. Демонстрирует значительное понимание проблемы. Большинство требований, предъявляемых к заданию, выполнены.
Отметкой (2-4 баллов) оценивается ответ, при котором студент в основном не правильно отвечает на поставленные вопросы. Демонстрирует частичное или небольшое понимание проблемы. Многие требования, предъявляемые к заданию, не выполнены.
Отметкой (0 -1 балл) оценивается ответ, при котором студент демонстрирует непонимание проблемы или нет ответа и даже не было попытки решить задачи.

В экзаменационный билет включены два теоретических вопроса и практическое задание, соответствующие содержанию формируемых компетенций. Зачет с оценкой проводится в письменной форме. На ответ и решение задачи студенту отводится 80 минут. За ответ на теоретические вопросы студент может получить максимально 15 баллов, за выполнение практических заданий - 15 баллов.

По итогам прохождения дисциплины и с учетом шкал оценивания все набранные в результате текущей, рубежной и промежуточной аттестаций баллы суммируются и выставляется оценка .

Перевод баллов в оценку:

85 - 100 баллов – «отлично»

70 - 84 баллов – «хорошо»

60 - 69 баллов – «удовлетворительно»

менее 60 баллов – «неудовлетворительно»

Раздел 5. Методические указания для обучающегося по освоению дисциплины и выполнению контрольных заданий

Рекомендации по организации самостоятельной работы студента

1. Советы по планированию и организации времени, необходимого для изучения дисциплины. Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины:

Изучение конспекта лекции в тот же день, после лекции – 10-15 минут;

Изучение конспекта лекции за день перед следующей лекцией – 10-15 минут;

Изучение теоретического материала по учебному пособию и конспекту – 1 час в неделю;

Подготовка к практическому занятию – 2-3 час.

Всего в неделю – 4 часа.

2. Описание последовательности действий студента

Для понимания материала и качественного его усвоения рекомендуется следующая последовательность действий.

После прослушивания лекции и окончания учебных занятий, при подготовке к занятиям следующего дня, нужно сначала просмотреть и обдумать текст лекции, прослушанной сегодня (10-15 минут).

При подготовке к лекции следующего дня, нужно просмотреть текст предыдущей лекции, подумать о том, какая может быть тема следующей лекции (10-15 минут).

В течение недели выбрать время (2-3часа) для работы с рекомендуемыми электронными учебными пособиями.

При подготовке к практическим занятиям следующего дня, необходимо сначала прочитать основные понятия и подходы по теме домашнего задания. При выполнении упражнения или задачи нужно сначала понять, что требуется в задаче, какой теоретический материал нужно использовать, наметить план решения задачи.

3. Рекомендации по использованию материалов учебно-методического комплекса. Все рекомендуемые учебные пособия размещены на серверах компьютерных классов в сетевой папке.