

**МИНИСТЕРСТВО ФИНАНСОВ РЕСПУБЛИКИ ТАДЖИКИСТАН**  
**ТАДЖИКСКИЙ ГОСУДАРСТВЕННЫЙ ФИНАНСОВО-**  
**ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ**  
**ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ЭКОНОМИКЕ**  
**КАФЕДРА ИНФОРМАЦИОННОГО И МАТЕМАТИЧЕСКОГО**  
**МОДЕЛИРОВАНИЯ**



Co-funded by the  
Erasmus+ Programme  
of the European Union

**Тест**  
**по предмету**  
**“АНАЛИЗ БОЛЬШИХ ДАННЫХ И**  
**БИЗНЕС-АНАЛИЗ”**

**Душанбе – 2021**

## ТЕСТЫ ПО ПРЕДМЕТУ «АНАЛИЗ БОЛЬШИХ ДАННЫХ И БИЗНЕС-АНАЛИЗ»

### 1. Что такое бизнес-процесс?

- A) Любая деятельность в корпоративных масштабах
- B) Коммерческая деятельность с целью получения прибыли
- C) Совокупность бизнес-функций
- D) Последовательность действий по преобразованию входов в выходы, удовлетворяющие потребителя

**2. Описать структуру системы бизнес-процессов, показать состав процессов одного уровня абстракции и взаимосвязи между ними можно с помощью диаграммы в нотации**

- A) EPC
- B) IDEF0
- C) BPMN
- D) DFD

### 3. Архитектура предприятия — это

- A) Искусство проектировать и строить бизнес-центры и производственные здания
- B) Концептуальная структура организация системы
- C) Единая система, которая описывает существующие организационные структуры, цели и показатели их достижения, линейку создаваемых продуктов/услуг, которые приносят доход, а также инфраструктуру (программное и аппаратное обеспечение, оборудование), используемые в работе
- D) Стил управления

**4. Требование «Пользовательский GUI должен предоставлять возможность языковой локализации: выбор языка**

**(русский/английский) для надписей на элементах» — это**

- A) Требование стейкхолдера (stakeholder requirement)
- B) Нефункциональное требование (non-functional requirement)
- C) Бизнес-требование (business requirement)
- D) Функциональное требование (functional requirement)

### 5. Владелец бизнес-процесса — это

- A) ответственный исполнитель
- B) лицо, которое отвечает за результат процесса, заинтересовано в нем, обладает ресурсами и полномочиями для его выполнения
- C) функциональный менеджер
- D) спонсор проекта

### 6. Аналог BPMN-диаграммы в UML — это

- A) **Диаграмма деятельности (activity diagram)**
- B) Диаграмма компонентов (Component diagram)
- C) Диаграмма классов (Class diagram)
- D) Диаграмма состояний (State diagram)

**7. Ключевым отличием проекта от процесса является**

- A) Требования к качеству результата
- B) Ограничение в ресурсах
- C) Обязательное наличие результата
- D) **Уникальность**

**8. Разработка требований к программному продукту в Agile-проектах характеризуется**

- A) нестабильным характером требований
- B) **итеративностью циклов детализации требований**
- C) появлением новых бизнес-потребностей
- D) отсутствием ТЗ (технического задания) по ГОСТ

**9. Диаграмма Исикавы (рыбья кость) нужна, чтобы**

- A) показать причинно-следственную связь процессов с результатом
- B) **определить потенциальные источники проблемы и оценить степень их влияния на результат**
- C) повысить уровень управляемости бизнес-процессов
- D) сформировать полный комплект документации СМК

**10. Организационная структура, которая предполагает двойное подчинение, например, начальнику функциональному отдела и менеджеру проекта, называется**

- A) Функциональная
- B) **Проектная**
- C) Процессная
- D) Распределенная

**11. Формат Parquet считается**

- A) неструктурированным
- B) полуструктурированным
- C) строковым
- D) **колоночным (столбцовым)**

**12. Для машинного обучения подходят данные**

- A) Любых форматов в цифровом виде
- B) Числовые типа int
- C) **Бинарные**

**D) Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов**

**13. Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД**

- A) HBase
- B) Cassandra
- C) Hive
- D) Elasticsearch**

**14. Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк**

- A) TensorFlow
- B) Flask
- C) PyTorch**
- D) Scikit-learn

**15. Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит**

- A) Apache Kafka**
- B) Apache Spark
- C) Apache AirFlow
- D) Apache Hadoop

**16. Apache NiFi используется для**

- A) визуализации результатов аналитики
- B) эффективного хранения больших данных
- C) маршрутизации потоков Big Data и построения ETL-конвейеров**
- D) оптимизации SQL-запросов к DWH

**17. Повысить производительность Apache Kafka можно с помощью:**

- A) Увеличения плотности разделов на каждом брокере
- B) Повышения коэффициента репликации
- C) Увеличения размера сообщений
- D) Замены HDD-дисков на SSD**

**18. Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью**

- A) Apache Hive
- B) Apache Hadoop
- C) Apache Kafka
- D) Apache AirFlow**

**19. Выберите технологию потоковой обработки событий в режиме реального времени**

- A) Spark Streaming
- B) Apache Kafka**
- C) Apache Hadoop
- D) MapReduce

**20. Объём накопленных человечеством цифровых данных на 2012 год измеряется:**

- A) петабайтами
- B) зеттабайтами**
- C) экзабайтами
- D) йоттабайтами

**21. Укажите фактор, способствовавший появлению тренда больших данных**

- A) маркетинговые кампании крупных корпораций**
- B) снижение издержек на хранение данных**
- C) появление новых технологий обработки потоковых данных
- D) выпуск баз данных с обработкой данных в памяти

**22. Какие вероятные разочарования тренда больших данных?**

- A) из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных**
- B) из-за угрозы безопасности личной жизни (privacy) граждан будут упрощены процедуры сбора данных, что приведёт к падению ценности больших данных**
- C) выпуск баз данных с обработкой данных в памяти
- D) нет

**23. Отметьте значимые события, повлиявшие на формирование тренда больших данных:**

- A) разработка Hadoop**
- B) изобретение принципа MapReduce**
- C) разработка языка Python
- D) победа Deepblue в матче с Г.Каспаровым

**24. Выберите верный ответ**

- A) большие данные – это обработка или хранение более 1 Тб информации
- B) проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна**

С) большие данные – это огромная PR-акция крупных вендоров и не более того

Д) большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект

**25. Выберите неверный ответ:**

А) большие данные – это данные объёма свыше 1 Тб

В) проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна

С) большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров

Д) большие данные как правило не структурированы

**26. Отметьте те из вариантов, в которых данные структурированы:**

А) данные о продажах компании, представленные в виде помесечных отчётов в формате MS Word

В) таблица с ежедневными показаниями температуры помещения за год в файле формата csv

С) текст педагогической поэмы А.С. Макаренко, представленный в формате PDF

Д) библиотека фильмов, представленных в формате mpeg4 на одном жестком диске

**27. Перечислите четыре основных характеристики Big Data:**

А) Virtualization, Volume, Variability, Vehicle

В) Variety, Velocity, Volume, Value

С) Verification, Volume, Velocity, Visualization

Д) Video, Value, Variety, Volume

**28. Выберите неверное высказывание:**

А) большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных

В) увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации

С) удешевление систем хранения на единицу информации привело к росту рынка больших данных

Д) разные типы данных в колонках таблиц реляционных СУБД

**29. Отметьте неверное понимание Variety в контексте характеристик Big Data:**

А) высокая скорость генерирования данных

В) разные типы данных в колонках таблиц реляционных СУБД

С) разнообразие отраслей, являющихся источниками данных

D) разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные

**30. Принцип MapReduce состоит в том, чтобы**

A) производить вычисления на узлах, где информация изначально была сохранена

B) использовать вычислительные мощности систем хранения

C) использовать функциональное программирование для решения задач массивно-параллельной обработки

D) нет правильного ответа

**31. Выберите одно неверное высказывание про MapReduce:**

A) интерфейс для массово-параллельной обработки данных, где вычисления производятся на узлах, где информация изначально была сохранена

B) MapReduce – это две операции: распределения и сборки данных

C) MapReduce был придуман разработчиками Hadoop

D) MapReduce был анонсирован разработчиками Google

**32. Какие из следующих технологий СУБД не используют принцип MapReduce**

A) Hadoop

B) Cassandra

C) HDInsight

D) Redis

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

A) Oracle Exalytics

B) SAP HANA

C) BigTable

D) HBase

**34. В чём преимущество колоночно-ориентированных СУБД?**

A) они позволяют выполнять более сложные SQL-запросы по сравнению с реляционными СУБД

B) они позволяют динамически дополнять содержание записей новыми полями

C) они имеют более гибкие возможности аналитики

D) они позволяют эффективно делать межколоночные сравнения

**35. Для чего аналитику необходима "песочница"?**

A) для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций

B) для хранения всех полученных от заказчика данных

C) для построения отчётов о результатах анализа

D) для снижения затрат, связанных с репликацией данных

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

A) Hadoop

B) Data Warehouse

C) "Песочница"

D) Python

**37. Выберите верное утверждение:**

A) Data Warehouse создаются для проверки гипотез при анализе больших данных

B) "Песочница" используется для снижения нагрузки на основной Data Warehouse

C) каждый Data Warehouse должен содержать "песочницу"

D) "Песочница" необходима для любого процесса аналитики

**38. Ниже приведена последовательность этапов проекта аналитики в соответствии с CRISP-DM, укажите первый этап.**

A) моделирование (Modeling)

B) внедрение (Deployment)

C) подготовка данных (Data Preparation)

D) понимание бизнеса (Business understanding)

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

A) понимание бизнеса (Business understanding)

B) понимание данных (Data Understanding)

C) моделирование (Modeling)

D) оценка (Evaluation)

**40. 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:**

A) понимание бизнеса (Business understanding)

B) подготовка данных (Data Preparation)

C) моделирование (Modeling)

D) оценка (Evaluation)

**41. Пример благоразумного использования Hadoop**

A) анализ 10 Гб данных

B) ежедневное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт)



- C) **посекундное сохранение данных температуры, поступающих со всех городов России (по одному показанию на город, всего городов 1100 шт)**
- D) **построение графика пульса пациента в реальном времени**

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

- A) 100Гб
- B) 1Тб
- C) **100Тб**
- D) 1Пб

**43. Hadoop – это:**

- A) **набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах**
- B) **распределённая СУБД, позволяющая обрабатывать большие данные**
- C) **язык выполнения заданий в парадигме MapReduce**
- D) **распределённая файловая система, предназначенная для хранения файлов большого объёма**

**44. Клиент покупает билет на самолет через интернет. В момент покупки, он хочет знать насколько может упасть стоимость этого билета в ближайшем будущем и когда. К какому типу относится эта задача анализа данных?**

- A) **прогнозирование**
- B) кластеризация
- C) классификация
- D) цензурирование

**45. Клиент покупает билет на самолет через интернет. Как бы в данном случае формулировалась задача прогнозирования?**

- A) **в момент покупки предсказать, насколько и когда может упасть стоимость этого билета в ближайшем будущем**
- B) **в момент покупки определить, к какому сегменту относится пользователь и предложить выгодные для него условия сделки**
- C) **выдать пользователю прогноз погоды для того места, куда он летит**
- D) **спрогнозировать вероятность дополнительных покупок (отель, страховка, туристический тур) и предложить наиболее вероятные пользователю**

**46. Клиент покупает билет на самолет через интернет. В момент покупки стоит задача определить вероятность дополнительных покупок (отель, страховка, туристический тур) и предложить наиболее вероятные пользователю. К какому типу относятся эта задача анализа данных?**

- A) заполнение пробелов
- B) классификация
- C) **прогнозирование**
- D) цензурирование

**47. Инвестиционный фонд интересуется тем, почему часть финансируемых им проектов успешно переходят на второй год, а часть - нет. К какому типу относится эта задача анализа данных?**

- A) **поиск информативных признаков**
- B) построение решающего правила
- C) классификация
- D) цензурирование

**48. Инвестиционный фонд имеет ряд проектов, который успешно переходят на второй год финансирования и тех, кто не переходит. Как бы в данном случае формулировалась задача поиска информативных признаков?**

- A) **определить, почему ряд проектов успешно переходят на второй год, а ряд – нет**
- B) определить для нового проекта, перейдёт ли он через год на второй этап финансирования или нет
- C) восстановить некоторые характеристики проектов, которые изначально не заполнялись
- D) **определить критерий успешности**

**49. Инвестиционный фонд имеет ряд проектов, который успешно переходят на второй год финансирования и тех, кто не переходит. Фонд поставил задачу определить критерий успешности проекта. К какому типу задач анализа данных наиболее близка эта задача?**

- A) прогнозирование
- B) построение решающего правила
- C) **поиск информативных признаков**
- D) цензурирование

**50. Коммерческая клиника желает установить структуру своих клиентов с точки зрения вклада в доход клиники. К какому типу относится эта задача анализа данных?**

- A) прогнозирование
- B) **кластеризация**
- C) классификация
- D) цензурирование

**51. Коммерческая клиника обладает некоторыми данными о клиентах и о доходах от них. Как бы в данном случае формулировалась задача кластеризации?**

- A) определить основные сегменты клиентов**
- B) определить, сколько денег принесёт тот или иной клиент в следующем периоде**
- C) определить, когда вернётся тот или иной клиент**
- D) определить, к какой группе "инноватор"- "последователь"- "консерватор" относится данный клиент**

**52. Коммерческая клиника обладает некоторыми данными о клиентах и о доходах от них. Руководителем поставлена задача определить, к какой группе "инноватор"- "последователь"- "консерватор" относится данный клиент. К каким из типов задач анализа данных близка эта задача?**

- A) кластеризация**
- B) прогнозирование**
- C) классификация**
- D) (4) таксономия**

**53. Компания, проводящая социологические опросы, испытывает сложности с верификацией данных, поступающих от волонтеров непосредственно опрашивающих респондентов: многие анкеты заполнены не полностью; волонтеры фальсифицируют результаты опроса, самостоятельно заполняя часть анкет. К какому типу наиболее близка эта задача анализа данных?**

- A) прогнозирование**
- B) кластеризация**
- C) классификация**
- D) цензурирование**

**54. Компания, проводящая социологические опросы получает анкеты от волонтеров, непосредственно опрашивающих респондентов. При каких условиях разумна постановка задачи цензурирования?**

- A) многие анкеты заполнены не полностью**
- B) стало известно, что волонтеры фальсифицируют результаты опроса, самостоятельно заполняя часть анкет**
- C) от заказчика поступило требование уничтожить часть анкет, содержащих информацию о руководителях страны**
- D) (4) Часть анкет пришла в негодность, что не позволяет считать информацию с них со 100% уверенностью**

**55. Компания, проводящая социологические опросы, испытывает сложности с верификацией данных, поступающих от волонтеров**

**непосредственно опрашивающих респондентов: многие анкеты заполнены не полностью; волонтеры фальсифицируют результаты опроса, самостоятельно заполняя часть анкет. К какому типу задач анализа данных здесь прибегать не придётся?**

- A) заполнение пробелов
- B) цензурирование
- C) прогнозирование
- D) классификация

**56. С некоторой периодичностью персонал предприятия списывает группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Руководство компании заинтересовано в сокращении количества проверок, при сохранении точности выявления ошибочного списания на уровне 97%. Требуется выявлять сомнительные акты списания, подлежащие обязательной проверке аудитором. К какому типу относится эта задача анализа данных?**

- A) прогнозирование
- B) кластеризация
- C) классификация
- D) цензурирование

**57. С некоторой периодичностью персонал предприятия списывает группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Как бы в данном случае формулировалась задача классификации?**

- A) определить характерные признаки ошибочных списаний
- B) научиться автоматически выявлять ошибочные списания с ожидаемой ошибкой не ниже 97%
- C) (3) классифицировать типичные ошибки и составить их список
- D) определить три категории: "ошибочные", "под сомнением", "безошибочные" и найти правило отнесения к этим категориям

**58. С некоторой периодичностью персонал предприятия списывает группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Определены три категории: "ошибочные", "под сомнением", "безошибочные". К какому типу задач анализа данных относится задача о построении правила автоматического отнесения списаний к этим категориям.**

- A) поиск информативных признаков
- B) кластеризация

С) классификация

Д) цензурирование

**59. Кардиологический центр исследовал возможность прогнозирования возникновения осложнений во время операционного вмешательства по до-операционным данным о пациенте: результатам физиологического и диагностического обследований. К какому типу задача анализа данных относится задача прогнозирования операционных осложнений?**

А) поиск информативных признаков

В) кластеризация

С) классификация

Д) цензурирование

**60. В кардиологическом центре проводятся операции. До и во время операционного вмешательства с пациента снимаются определённые показания. Известны случаи осложнений во время проведения операций, эти случаи описаны тем же набором данных. Как бы в данном случае формулировалась задача классификации?**

А) сгруппировать пациентов по похожим дооперационным показаниям

В) спрогнозировать по дооперационным данным вероятность возникновения осложнений во время операции

С) ранее были определены три группы риска среди пациентов. Все предыдущие случаи осложнений отнесены к определённым группам риска. Стоит задача определить группу риска для нового пациента, прошедшего дооперационные обследования

Д) часть данных о пациентах не содержит некоторых полей, требуется заполнить эти данные на основе статистики, чтобы включить данные об этих пациентах в выборку

**61. В кардиологическом центре проводятся операции. До и во время операционного вмешательства с пациента снимаются определённые показания. Известны случаи осложнений во время проведения операций, эти случаи описаны тем же набором данных. Ранее были определены три группы риска среди пациентов. Все предыдущие случаи осложнений отнесены к определённым группам риска. Стоит задача определить группу риска для нового пациента, прошедшего дооперационные обследования. К какому типу относится эта задача анализа данных?**

А) прогнозирование

В) кластеризация

С) классификация

Д) цензурирование

**62. К какому типу шкал относится шкала "очень плохо"-"плохо"-"средне"-"хорошо"-"очень хорошо"?**

- A) порядковая
- B) абсолютная
- C) бинарная
- D) номинальная

**63. Какие из следующих шкал в точности являются порядковыми?**

- A) "Слабо", "Сильно"
- B) "Красный", "Желтый", "Зеленый"
- C) "удовлетворительно", "хорошо", "отлично"
- D) показания термометра

**64. На основе какого языка был создан R?**

- A) C
- B) S
- C) Java
- D) Python

**65. Какие из перечисленных языков оказали влияние на R?**

- A) Ruby
- B) Matlab
- C) Scheme
- D) S

**66. R является:**

- A) языком программирования для статистической обработки данных и работы с графикой, созданный на основе языка S
- B) свободной программной средой вычислений с открытым исходным кодом
- C) высокоуровневым языком программирования общего назначения, ориентированным на повышение производительности разработчика и читаемости кода
- D) языком программирования с динамической типизацией данных

**67. Присваивание в языке R можно осуществить с помощью оператора:**

- A) =
- B) <-
- C) ->
- D) :=

**68. Присваивание в R нельзя осуществить с помощью оператора:**

- A) =
- B) <-
- C) ->

D) :=

**67. Какие из вариантов присвоения не вызовут ошибки:**

- A)  $x \rightarrow 3$
- B)  $x <- 3$
- C)  $3 <- x$
- D)  $3 \rightarrow x$

**68. Как сделать вектор из трех чисел?**

- A) (7,7,7)
- B) [7,7,7]
- C) {7,7,7}
- D) c(7,7,7)

**69. В каких из следующих случаях исполнение выдаст ошибку?**

- A) (7,7,7)
- B) [7,7,7]
- C) {7,7,7}
- D) c(7,7,7)

**70. Для каких аргументов функция is.finite вернет true?**

- A) 1
- B) NA
- C) NaN
- D) (+Inf)

**71. Для каких аргументов функция is.finite вернет false?**

- A) 1
- B) NA
- C) NaN
- D) (+Inf)

**72. Какой результат будет у функции is.finite(Inf-Inf)?**

- A) TRUE
- B) FALSE
- C) NAN
- D) ошибка

**73. Что вернет следующее выражение sum(1:3>2)?**

- A) 0
- B) 1
- C) 2
- D) 3. Ошибка

**74. Каким образом можно сформировать вектор (FALSE, FALSE, TRUE)?**

- A) c(FALSE, FALSE, TRUE)

B)  $2:4>3$

C)  $2:4<3$

D)  $-c(\text{TRUE}, \text{TRUE}, \text{FALSE})$

**75.  $p.x=1.y=2$  Что выведет  $\text{print}(p)$ ?**

A) (1,2)

B)  $\{x=1;y=2\}$

C) NaN

D) ошибка

**76. Как правильно присвоить  $p$  значение типа данных, имеющих две координаты  $x=1$  и  $y=2$ ?**

A)  $p.x=1.y=2$

B)  $p = c(x=1, y=2)$

C)  $p = c(y=2, x=1)$

D)  $p(x,y)<-c(1,2)$

**77. Какая типизация в языке R?**

A) статическая

B) динамическая

C) статическая с элементами динамической

D) типизация отсутствует

**78. Выберите неверные утверждения:**

A) язык R имеет статическую типизацию

B) язык R имеет динамическую типизацию

C) язык R имеет статическую типизацию с элементами динамической

D) язык R не имеет типизации

**79. Операция с диапазонами  $x=1:3$ ,  $y=4:6$ . Чему будет равно  $x+y$ ?**

A) 5 7 9

B) 1 2 3 4 5 6

C) 5:09

D) 1:06

**80. Каким образом можно описать вектор (1,2,3,4,5,6)?**

A)  $1:2+4:6$

B)  $1:8-7:8$

C) 1:06

D)  $2:12/2$

**81. Есть таблица со столбцами  $x$ ,  $id$ ,  $y$ ,  $price$ . Какая формула описывает зависимость  $price$  от  $x,y$ ?**

A)  $price \sim x + y$

B)  $price \sim . - id$

C)  $x, y \sim price$



D) *price* ~

**82. Каким будет округленный результат прогнозирования пустого значения в таблице с помощью функции ТЕНДЕНЦИЯ:**

Год        1   2   3   4

Тыс. руб 67 120 160

A) **209**

B) 210

C) 220

D) 301

**83. Какое значение запишет функция Flash (Мгновенное заполнение) Excel в пропущенной ячейке:**

Название	Жанр
Курочка Ряба	Сказка
Дюймовочка	Сказка
Война и мир	

A) война и мир

B) роман

C) **сказка**

D) Фильм

**84. Существуют ли встроенные ограничения в Excel 2013 на количество элементов данных в одном ряду данных для плоских диаграмм?**

A) **да, не более 32 000**

B) да, не более 32 767

C) да, не более 1 048 576

D) нет, размер ограничен только объемом доступной оперативной памяти

**85. Каково ограничение в Excel 2013 на количество строк при загрузке csv-файла?**

A) **1048576**

B) 65535

C) нет, размер ограничен только объемом доступной оперативной памяти

D) 2147483648

**86. Отметьте верное представление типа данных "очень плохо"- "плохо"- "средне"- "хорошо"- "очень хорошо" в файле типа ARFF:**

A) **@attribute estimation { “очень плохо”, плохо, средне, хорошо, “очень хорошо”}**

B) @attribute estimation { очень плохо, плохо, средне, хорошо, очень хорошо}

C) @attribute estimation { [очень ]плохо, средне, [очень] хорошо}

D) @attribute estimation { “[очень ]плохо”, средне, “[очень ]хорошо”}

**87. Какие из перечисленных признаков таблицы могут быть классификационными без предварительной обработки?**

Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

A) Temperature, Humidity

B) **Play, Outlook, Wind**

C) Play, Outlook, Temperature, Humidity, Wind

D) Row No., Play, Outlook, Temperature, Humidity, Wind

**88. Какие из перечисленных признаков таблицы представлены в абсолютной шкале?**

Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

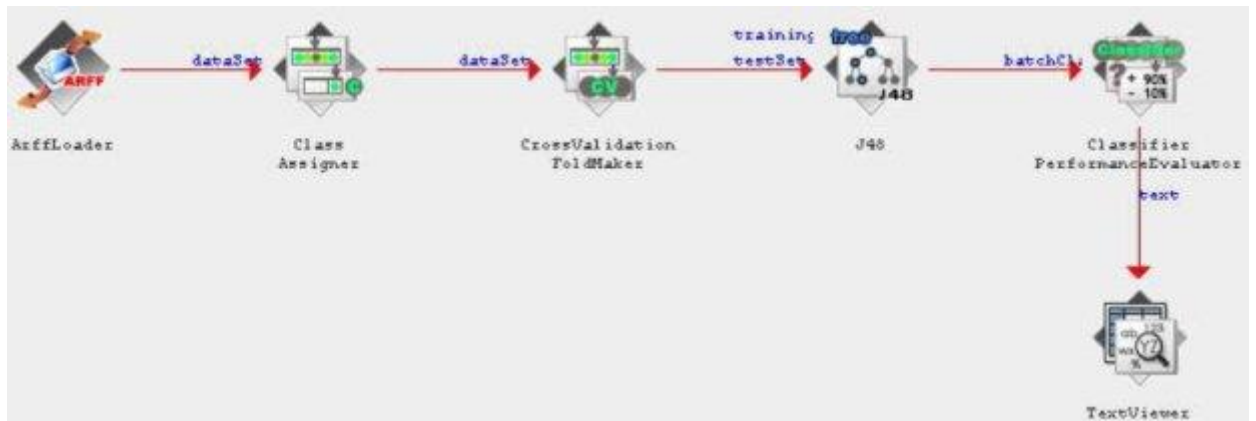
A) **никакие**

B) Temperature

C) Humidity

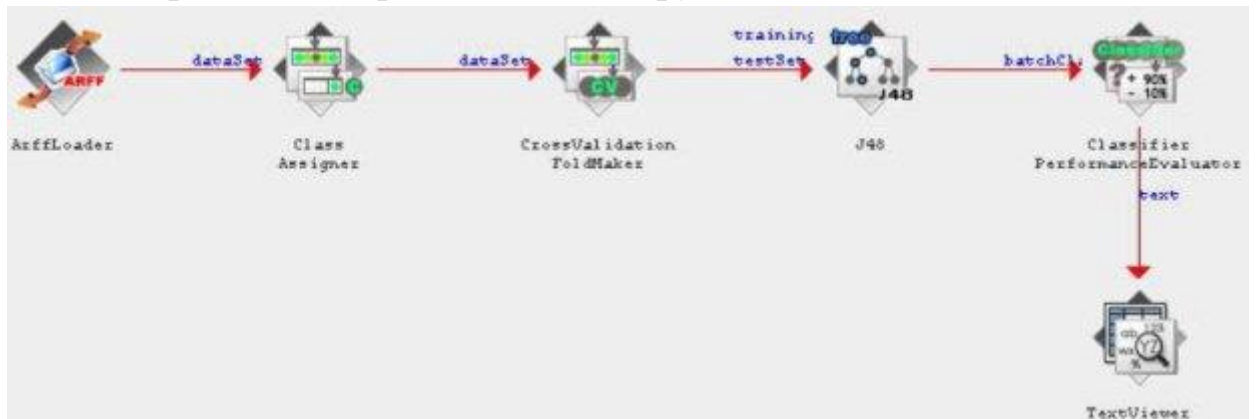
D) Temperature, Humidity

**89. сновной целью процесса на приведенной диаграмме является**



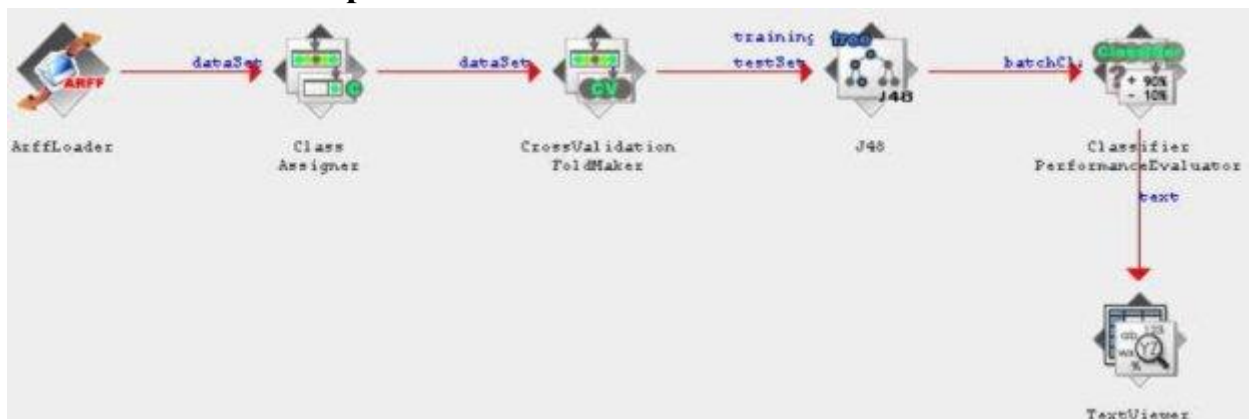
- A) произвести оценку качества работы алгоритма J48
- B) классифицировать выборку объектов алгоритмом J48
- C) оценить качество исходных данных
- D) выбрать наиболее релевантную обучающую выборку

**90. Можно ли с помощью приведённой схемы произвести оценку качества работы алгоритма J48 на загруженных данных?**



- A) нет
- B) да, достаточно одного прогона
- C) да, только после 10 запусков
- D) да, только после 100 запусков

**91. Что выведет в приведённой схеме TextViewer?**



- A) число
- B) текст

- C) обучающую выборку
- D) контрольную и обучающую выборку

**92. Продукты аналогичные Knime, RapidMiner хорошо подойдут для**

- A) быстрого прототипирования модели обработки данных
- B) создания гибкой интерактивной модели обработки данных
- C) визуализации данных
- D) для разработки новых алгоритмов

**93. В каком случае применение Tableau наиболее оправдано**

- A) проведено исследование, результатом которого стала таблица объект-свойства, необходимо предоставить отчетность
- B) имеются данные, необходимо более получить ясное понимание этих данных
- C) необходимо реализовать гибкое интерактивное визуальное представление данных
- D) не оправдано

**94. Какой из языков наиболее богат библиотеками/фреймворками для интерактивной визуализации данных**

- A) **Java Script**
- B) R
- C) Python
- D) c

**95. Какая из приведённых функций реализует алгоритм классификации**

- A) kmeans()
- B) hcclust()
- C) **randomForest()**
- D) classify()

**96. Какая из функций реализует алгоритм кластеризации**

- A) **pam()**
- B) nnet()
- C) party()
- D) cluster()

**97. Алгоритм какого типа реализует функция randomForest()?**

- A) кластеризации
- B) **классификации**
- C) регрессии
- D) визуализации в виде деревьев

**98. Основное преимущество Python перед R – это**

- A) **быстродействие**
- B) элементы функционального программирования

C) Python превосходит R, как язык программирования

D) строго типизирован

**99. Назовите преимущества R, как инструмента data mining, перед другими языками**

A) быстродействие

B) большое кол-во библиотек алгоритмов ИАД и статистических пакетов

C) удобная работа с данными

D) работа с большими данными

**100. базовая библиотека для data science на Python**

A) Scipy stack (NumPy & SciPy)

B) Theano

C) MDP

D) Mahout