

«УТВЕРЖДАЮ»
Декан факультета
компьютерной инженерии
_____ Т. Кучкоров
«___» _____ 2026 г.

**Министерство цифровых технологий Республики Узбекистан
Ташкентский университет информационных технологий имени
Мухаммада аль-Хорезми**

**ВОПРОСЫ ПО ПРЕДМЕТУ “DATA MINING”
ДЛЯ ПРОВЕДЕНИЯ ИК**

1. Что понимается под интеллектуальным анализом данных (Data Mining)?
2. В чем заключается основная цель интеллектуальной обработки данных?
3. Какие задачи решает интеллектуальный анализ данных?
4. Почему возникла необходимость в развитии Data Mining?
5. Какие этапы прошёл анализ данных в процессе своего развития?
6. В чем отличие традиционного анализа данных от Data Mining?
7. Что такое природные (неструктурированные) данные? Приведите примеры.
8. Что относится к полуструктурированным данным?
9. Что понимается под сложными данными и какие их особенности?
10. Какие трудности возникают при работе с неструктурированными и сложными данными?
11. Какие основные области применяются в интеллектуальном анализе данных (Data Mining)?
12. Как связаны интеллектуальный анализ данных и статистика?
13. В чем заключается связь Data Mining с машинным обучением?
14. Какую роль играют базы данных в интеллектуальном анализе данных?
15. Как используется искусственный интеллект в Data Mining?
16. Как теория вероятностей применяется в интеллектуальном анализе данных?
17. Какова роль информатики в развитии Data Mining?
18. Почему Data Mining считается мультидисциплинарной областью?
19. Какие преимущества даёт объединение разных научных областей в Data Mining?
20. Приведите примеры задач, где используются методы из разных дисциплин в Data Mining.
21. Какие существуют среды сбора и хранения данных?
22. Какие основные источники данных используются в Data Mining?
23. Какую роль играют датчики (sensors) в сборе данных?

24. Что представляют собой журналы (logs) как источник данных?
25. Что такое транзакционные данные и где они применяются?
26. Что такое СУБД и какие функции она выполняет?
27. В чем отличие реляционных баз данных от NoSQL систем?
28. Что такое Apache Hadoop и для чего он используется?
29. В чем разница между сбором данных в реальном времени и пакетной обработкой?
30. Какие преимущества и недостатки имеют методы сбора данных в реальном времени и в пакетном режиме?
31. Что такое предварительная обработка данных и зачем она необходима?
32. Какие основные этапы включает процесс предварительной обработки данных?
33. Что такое очистка данных (data cleaning)?
34. Какие методы используются для обнаружения аномалий в данных?
35. Что такое пропущенные значения и какие существуют способы их обработки?
36. В чем разница между удалением и заполнением пропусков?
37. Что такое кодирование категориальных признаков и какие методы существуют?
38. Что такое нормализация данных и зачем она применяется?
39. Что включает в себя трансформация признаков?
40. Что такое отбор признаков (feature selection) и почему он важен?
41. Что такое интеграция данных и зачем она необходима?
42. Какие проблемы возникают при объединении данных из разных источников?
43. Какие методы используются для интеграции многоисточниковых данных?
44. Что такое процесс ETL (Extract, Transform, Load)?
45. Какие этапы включает ETL-процесс?
46. Что понимается под концепцией data pipeline?
47. В чем отличие ETL от ELT?
48. Какие методы используются для сжатия данных?
49. Как сжатие данных влияет на производительность и использование памяти?
50. Какие преимущества дает эффективное управление памятью при работе с большими данными?
51. Что такое хранилище данных (Data Warehouse)?
52. Какие основные характеристики отличают хранилище данных от обычных баз данных?
53. В чем заключается разница между системами OLTP и OLAP?
54. Для каких задач используются OLTP-системы?
55. Какие задачи решают OLAP-системы?
56. Что включает архитектура хранилища данных?
57. Что такое ODS (Operational Data Store) и какую роль он выполняет?

58. Что представляет собой staging area в хранилище данных?
59. Что такое data mart и чем он отличается от Data Warehouse?
60. В чем разница между операционной и аналитической обработкой данных?
61. Что такое OLAP-технология и где она применяется?
62. В чем заключается суть многомерного анализа данных?
63. Что такое куб данных (data cube)?
64. Какие элементы входят в модель многомерных данных (измерения, факты)?
65. Что такое операция roll-up в OLAP?
66. В чем заключается операция drill-down (drill up)?
67. Что означает операция slice в OLAP?
68. Что такое операция dice и чем она отличается от slice?
69. Как выполняется моделирование измерений в OLAP-системах?
70. Как OLAP интегрируется с BI-системами и какие преимущества это дает?
71. Почему важно эффективно вычислять объемы данных в OLAP-системах?
72. Какие проблемы возникают при работе с большими многомерными кубами данных?
73. Какие методы используются для уменьшения размера куба данных?
74. В чем заключается идея агрегирования данных для сокращения объема?
75. Что такое Iceberg Cube и в каких случаях он применяется?
76. Что представляет собой lattice (решетчатая структура) куба данных?
77. Как взаимосвязаны Iceberg-модель и lattice в контексте анализа данных?
78. Что понимается под сходством данных и как оно измеряется?
79. Какие метрики используются для оценки сходства объектов?
80. Что такое ассоциативный анализ и где он применяется?
81. Что такое часто встречающиеся наборы элементов (frequent itemsets)?
82. Какие условия определяют частоту набора элементов?
83. Что такое ассоциативные правила?
84. Какие показатели используются для оценки ассоциативных правил (support, confidence, lift)?
85. Что такое корреляционный анализ и как он может быть визуализирован?
86. Что такое задача классификации и чем она отличается от задачи прогнозирования?
87. В чем заключается принцип работы деревьев решений (Decision Trees)?
88. Какие преимущества и недостатки имеют деревья решений?
89. Что такое метод опорных векторов (SVM) и где он применяется?
90. В чем особенность алгоритма k ближайших соседей (K-NN)?
91. Как влияет выбор параметра k на качество классификации в K-NN?

92. В чем заключается суть логистической регрессии и когда её используют?
93. Что такое линейная регрессия и как она применяется для прогнозирования?
94. В чем отличие полиномиальной регрессии от линейной?
95. Как оценивается качество моделей классификации и прогнозирования?
96. Что такое кластеризация и чем она отличается от классификации?
97. Какие преимущества имеет неконтролируемый характер кластеризации?
98. Как работает алгоритм K-means и какие шаги он выполняет?
99. Какие недостатки имеет алгоритм K-means?
100. В чем заключается суть иерархической кластеризации?
101. Что такое дендрограмма и как она используется в иерархической кластеризации?
102. Как работает алгоритм DBSCAN и чем он отличается от K-means?
103. Какие параметры важны для DBSCAN и как они влияют на результат?
104. Как определить оптимальное число кластеров при использовании K-means?
105. Какие методы визуализации применяются для анализа кластеров и оценки их качества?

Заведующий кафедрой

Б.Р. Азимов

Ответственный составитель

М.А. Узакова