**Экзаменационная вопросы по дисциплине**

**Интеллектуальный анализ данных**

1. Опишите основные задачи интеллектуального анализа данных, которые применяются в различных сферах деятельности.
2. Дайте понятие интеллектуального анализа данных и объясните, как оно связано с понятием больших данных.
3. Охарактеризуйте этапы процесса интеллектуального анализа данных и их роль в решении практических задач.
4. Перечислите и объясните основные методы интеллектуального анализа данных, такие как классификация, кластеризация и прогнозирование.
5. Раскройте суть понятий «обучение с учителем» и «обучение без учителя» в контексте интеллектуального анализа данных.
6. Опишите основные этапы поиска источников данных и их значимость для последующего анализа.
7. Дайте понятие первичных и вторичных данных и охарактеризуйте их особенности.
8. Охарактеризуйте методы сбора данных, такие как опросы, наблюдение и эксперименты, и объясните, в каких ситуациях каждый из них наиболее эффективен.
9. Перечислите и объясните критерии оценки надежности и достоверности источников данных.
10. Опишите, как современные технологии, такие как веб-скрейпинг и сенсоры, применяются для сбора данных в различных областях.
11. Дайте понятие стемминга и объясните его основную цель при обработке текстовых данных.
12. Охарактеризуйте разницу между стеммингом и лемматизацией, указав их преимущества и недостатки.
13. Опишите, как алгоритмы стемминга, такие как Porter Stemmer и Snowball Stemmer, работают с текстовыми данными.
14. Перечислите основные этапы выполнения операции стемминга и объясните, как они влияют на обработку текста.
15. Объясните, в каких задачах обработки текстов стемминг наиболее полезен, приведя примеры его практического применения.
16. Дайте определение технологии обработки естественных языков (NLP), опишите в каких областях она широко используется в настоящее время.
17. Опишите основные этапы и основные операции по обработке текстовых данных в области обработки естественных языков (NLP).
18. Опишите основные концепции контекстно-свободной грамматики и методов исчисления предикатов.
19. Дайте характеристику основным этапам обработки текстов в NLP, нарисуйте схему этапов обработки текстов.
20. Приведите характеристику различных методов векторизации текстов, таких как Bag of words, Tf-idf и другие.
21. Опишите векторизацию текстов с помощью word embeddings: Word2Vec, FastText и т.д.
22. Опишите основные этапы подготовки текстовых данных для машинного обучения и их важность для качества модели.
23. Дайте понятие токенизации и объясните, как она применяется при обработке текстовых данных.
24. Охарактеризуйте процесс нормализации текста, включая такие методы, как приведение к нижнему регистру и удаление пунктуации.
25. Перечислите и объясните методы представления текстовых данных в числовом формате, такие как Bag-of-Words, TF-IDF и векторизация.
26. Объясните, почему удаление стоп-слов и очистка данных от шумов являются ключевыми шагами в подготовке текстов.
27. Дайте понятие классификации в машинном обучении и объясните, как она отличается от других типов задач, таких как регрессия и кластеризация.
28. Охарактеризуйте разницу между бинарной и многоклассовой классификацией, приведя примеры задач для каждой.
29. Перечислите популярные алгоритмы классификации, такие как логистическая регрессия, деревья решений и метод опорных векторов, и опишите их особенности.
30. Объясните, что такое переобучение в задачах классификации, и предложите методы его предотвращения.
31. Опишите, как метрики оценки, такие как точность, полнота, F-мера и ROC-кривая, применяются для анализа производительности классификационных моделей.
32. Дайте понятие ансамблевых методов классификации, таких как Random Forest и Gradient Boosting, и объясните их преимущества.
33. Охарактеризуйте роль функции активации и архитектуры нейронной сети в задачах классификации.
34. Объясните, как методы обработки несбалансированных данных, такие как ресемплирование и модификация функции потерь, влияют на качество классификации.
35. Опишите использование методов кросс-валидации при обучении и оценке классификационных моделей, а также их роль в обеспечении стабильности результатов.
36. Опишите архитектуру рекуррентных нейронных сетей (RNN) и объясните, почему они подходят для работы с последовательностями текстов.
37. Охарактеризуйте роль векторизации текста, включая методы Word2Vec, GloVe и FastText, в задачах классификации текстовых данных.
38. Объясните, как двунаправленные RNN (BiRNN) и LSTM улучшают обработку текста по сравнению с обычными RNN.
39. Опишите использование сверточных нейронных сетей (CNN) для анализа текстовых данных и их преимущества в задачах классификации.
40. Дайте понятие трансформеров и объясните, как модели, такие как BERT и GPT, применяются для классификации текстов.
41. Напишите скрипт для обработки текстовых данных: токенизации, удаления стоп-слов и приведения к нижнему регистру.
42. Напишите скрипт для создания и обучения простой нейронной сети в Keras для задачи регрессии.
43. Напишите скрипт для выполнения кластеризации методом K-means и визуализации кластеров на графике.
44. Напишите скрипт для автоматического подбора гиперпараметров модели с использованием GridSearchCV.
45. Напишите скрипт для загрузки изображений из директории, их масштабирования и подготовки для ввода в сверточную нейронную сеть.
46. Напишите скрипт для оценки производительности модели классификации с помощью метрик, таких как точность, полнота и F-мера.
47. Напишите скрипт для применения PCA (метод главных компонент) для уменьшения размерности данных.
48. Напишите Keras скрипт для построения и обучения модели нейронной сети с одной скрытой слоем для задачи бинарной классификации.
49. Напишите Keras скрипт для реализации и обучения сверточной нейронной сети (CNN) на наборе данных CIFAR-10.
50. Напишите Keras скрипт для предобработки данных, включая нормализацию и разбиение на обучающую и тестовую выборки.
51. Напишите Keras скрипт для использования callback-функций LearningRateScheduler для динамической настройки скорости обучения.
52. Напишите Keras скрипт для построения LSTM-модели для прогнозирования временных рядов, включая визуализацию предсказаний.
53. Напишите скрипт для визуализации дерева решений, обученного на данных с помощью Scikit-learn.
54. Напишите скрипт для загрузки данных из файла CSV и их визуализации с использованием библиотеки Matplotlib.
55. Напишите скрипт для построения линейной регрессии с использованием библиотеки Scikit-learn и оценки её точности.
56. Дан небольшой скрипт. Дополни его команда для классификации с помощью Decision Tree.

import tensorflow as tf

from tensorflow.keras.preprocessing.text import Tokenizer

from tensorflow.keras.preprocessing.sequence import pad\_sequences

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense, Embedding, GlobalAveragePooling1D

# Пример данных

reviews = [

    "Great movie! Must watch.",

    "Worst movie ever.",

    "It was a fantastic performance!",

    "Not a good movie, boring plot.",

    "The cinematography is pretty great.",

    "I hated the storyline."

]

labels = [1, 0, 1, 0, 1, 0]  # 1 - положительный, 0 - отрицательный

1. Дан небольшой скрипт. Дополни его команда для классификации с помощью Random Forest.

import tensorflow as tf

from tensorflow.keras.preprocessing.text import Tokenizer

from tensorflow.keras.preprocessing.sequence import pad\_sequences

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense, Embedding, GlobalAveragePooling1D

# Пример данных

reviews = [

    "Great movie! Must watch.",

    "Worst movie ever.",

    "It was a fantastic performance!",

    "Not a good movie, boring plot.",

    "The cinematography is pretty great.",

    "I hated the storyline."

]

labels = [1, 0, 1, 0, 1, 0]  # 1 - положительный, 0 - отрицательный

1. Дан небольшой скрипт. Дополни его команда для классификации с помощью Dense neural network

import tensorflow as tf

from tensorflow.keras.preprocessing.text import Tokenizer

from tensorflow.keras.preprocessing.sequence import pad\_sequences

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense, Embedding, GlobalAveragePooling1D

# Пример данных

reviews = [

    "Great movie! Must watch.",

    "Worst movie ever.",

    "It was a fantastic performance!",

    "Not a good movie, boring plot.",

    "The cinematography is pretty great.",

    "I hated the storyline."

]

labels = [1, 0, 1, 0, 1, 0]  # 1 - положительный, 0 - отрицательный

1. Дан небольшой скрипт. Дополни его команда для классификации с помощью Long short-term memory neural network

import tensorflow as tf

from tensorflow.keras.preprocessing.text import Tokenizer

from tensorflow.keras.preprocessing.sequence import pad\_sequences

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense, Embedding, GlobalAveragePooling1D

# Пример данных

reviews = [

    "Great movie! Must watch.",

    "Worst movie ever.",

    "It was a fantastic performance!",

    "Not a good movie, boring plot.",

    "The cinematography is pretty great.",

    "I hated the storyline."

]

labels = [1, 0, 1, 0, 1, 0]  # 1 - положительный, 0 - отрицательный

1. Напишите скрипт по созданию программного кода на Keras для Глубоких нейронных сетей (Dense neural network)