

Оценочные материалы для промежуточной аттестации по дисциплине

Название дисциплины «Большие данные», 6 семестр

Код, направление подготовки	38.03.05 Бизнес-информатика
Направленность (профиль)	Экономика предприятий управление бизнес-процессами
Форма обучения	Очная
Кафедра-разработчик	Менеджмента и бизнеса
Выпускающая кафедра	Менеджмента и бизнеса

Типовые задания для контрольной работы:

1. Задачи и терминология больших данных: supervised и unsupervised задачи; регрессия, классификация, кластеризация.
2. Объект, признак, типы признаков, методы работы с ними. Метрики качества. Инструменты интеллектуального анализа данных.
3. Изучение основ работы с векторными данными и визуализацией. Градиент.
4. Методы оптимизации гладких функций.
5. Реализация градиентного спуска для линейной регрессии.
6. Выделение признаков из текста: one-hot encoding, стемминг, лемматизация, tf-idf преобразование.
7. Логистическая регрессия на примере задачи эмоциональной окраски текстов. L1 и L2 регуляризация.
8. Способы оценки качества моделей: holdout и кросс-валидация.
9. Метод ближайших соседей.
10. Деревья решений, случайный лес, градиентный бустинг.
11. Решение соревнования на платформе Kaggle.
12. Способы построения композиций моделей.
13. Алгоритмы кластеризации: K-means, DBscan, агломеративная кластеризация.
14. Методы понижения размерности на основе матричных разложений (PCA и SVD). T-SNE.
15. Введение в нейронные сети - полносвязные нейросети, метод обратного распространения ошибки, инициализация весов, нелинейности.
16. Обзор стохастических методов оптимизации первого порядка.

Можно предложить свой вариант темы

(согласовать с преподавателем обязательно)!

При оформлении контрольной работы необходимо соблюдать следующую структуру:

1. Титульный лист
2. Оглавление (сформированное автоматически)

3. Введение
4. Содержательная часть
(теория, **решение индивидуальной задачи или примера обязательно проработать самостоятельно**)
5. Заключение
6. Список литературы.

Вопросы для подготовки к зачету.

1. Задачи и терминология больших данных: supervised и unsupervised задачи; регрессия, классификация, кластеризация.
2. Объект, признак, типы признаков, методы работы с ними. Метрики качества. Инструменты интеллектуального анализа данных.
3. Изучение основ работы с векторными данными и визуализацией. Градиент.
4. Методы оптимизации гладких функций.
5. Реализация градиентного спуска для линейной регрессии.
6. Выделение признаков из текста: one-hot encoding, стемминг, лемматизация, tf-idf преобразование.
7. Логистическая регрессия на примере задачи эмоциональной окраски текстов. L1 и L2 регуляризация.
8. Способы оценки качества моделей: holdout и кросс-валидация.
9. Метод ближайших соседей.
10. Деревья решений, случайный лес, градиентный бустинг.
11. Решение соревнования на платформе Kaggle.
12. Способы построения композиций моделей.
13. Алгоритмы кластеризации: K-means, DBscan, агломеративная кластеризация.
14. Методы понижения размерности на основе матричных разложений (PCA и SVD). T-SNE.
15. Введение в нейронные сети - полносвязные нейросети, метод обратного распространения ошибки, инициализация весов, нелинейности.
16. Обзор стохастических методов оптимизации первого порядка.