**Экзаменационные вопросы по дисциплине «Методы машинного обучения и анализа данных на R»**

1. Генеральная совокупность и выборка. Типы переменных в статистике. Описательная статистика. Меры центральной тенденции. Меры изменчивости и их свойства. Квартили распределения
2. Нормальное распределение. Z-преобразование. Правило 2-х и 3-х сигм. Центральная предельная теорема. Доверительные интервалы для среднего
3. Идея статистического вывода. Статистический критерий. Свойства критериев. Алгоритм проверки статистической гипотезы. Параметрические и непараметрические критерии. Уровни значимости критерия (ошибки).
4. t-распределение. Сравнение двух средних. t-критерий Стьюдента. Сравнение двух средних. Непараметрический тест Манна-Уитни
5. Способы проверки распределения на нормальность: тесты Колмогорова-Смирнова и Шапиро-Уилка. Графические способы проверки распределения на нормальность: гистограмма и Q-Q plot.
6. Однофакторный дисперсионный анализ. Критерий Фишера. Множественные сравнения в ANOVA. Поправки Бонферрони и Тьюки
7. Многофакторный ANOVA. АБ-тестирование. Корреляция (коэффициенты Пирсона, Спирмена и Тау-Кендала)
8. Гипотеза о значимости взаимосвязи. Коэффициент детерминации. Метод наименьших квадратов. Применение регрессионного анализа.
9. Критерии согласия. Хи-квадрат. Колмогорова-Смирнова. Шапиро-Уилка. Лилиефорса. Критерий Бартлетта
10. Подготовка данных в задачах машинного обучения. Алгоритм градиентного спуска и его модификации
11. Задача классификации. Логистическая регрессия. Постановка задачи, входные данные и алгоритм. Задача регрессии. Линейные модели.
12. Переобучение и борьба с ним. Метрики качества классификации. Метрики качества регрессии
13. Метод K ближайших соседей. Постановка задачи кластеризации. Алгоритм k-means. Алгоритм DBSCAN
14. Решающие деревья. Случайные леса. Ансамблевые методы: стекинг, бэггинг, бустинг.
15. Задача понижения размерности. Отбор признаков. Выделение новых признаков на основе старых
16. Ассоциативные правила. Алгоритм Apriori. Рекомендательные системы