**Машинное обучение, анализ данных и нейронные сети в дизайне химических веществ с заданными свойствами**

**Machine learning, data analysis, and neural networks in the rational design of chemical compounds**

*Вопросы к зачету*

**Подготовьте презентацию** на 2-3 слайда (не включая заглавный), кратко иллюстрирующую принцип работы указанных ниже методов машинного обучения. Информация на слайде должна как минимум иллюстрировать материал курса по указанной теме, однако использование рисунков из презентаций преподавателя запрещается.

1. Линейная и логистическая регрессия.
2. Метод главных компонент (principal component analysis).
3. Алгоритм k-ближайших соседей (k-nearest neighbors) и метод опорных векторов (support vector machines).
4. Дерево решений и случайный лес (random forest).
5. Искусственные нейронные сети.

На зачете студенту также необходимо **кратко, устно** ответить на два случайных вопроса из списка ниже.

1. Какие форматы записи экспериментальных/расчетных данных используются в химических исследованиях (назовите минимум три)?
2. Назовите как минимум один формат записи 2D-структуры молекул и запишите структуру органической молекулы на ваш выбор, используя его.
3. Опишите структуру .xyz-файла. Для чего используется этот формат данных?
4. Опишите структуру MDL Mol-файла. Для чего используется этот формат данных?
5. Опишите структуру PDB-файла. Для чего используется этот формат данных?
6. Назовите как минимум две базы данных, содержащие информацию о химических реакциях? Какие проблемы для их использования в машинном обучении можно выделить?
7. Назовите как минимум два выражения для линейной зависимости свободных энергий. Какие свойства химических систем они связывают?
8. Приведите пример как минимум трех скалярных дескрипторов молекулярной структуры.
9. Приведите пример как минимум двух векторных дескрипторов молекулярной структуры.
10. Приведите пример тензорного дескриптора молекулярной структуры.
11. Какие физическо-химические свойства химических систем можно выбрать в качестве признака (feature) при рациональном дизайне медицинского препарата, катализатора, материала с помощью машинного обучения? Назовите минимум пять признаков.
12. Какие свойства могут быть ключевыми при разработке лекарственных препаратов? Назовите минимум два.
13. Что такое "электронный язык"? Можете ли вы объяснить принцип действия этого устройства?
14. Какие свойства катализатора могут быть ключевыми в рациональном дизайне? Назовите минимум два.
15. Приведите пример минимум трех задач искусственного интеллекта в органическом синтезе.