**Программа курса ИИС-2018**

1. Постановка задачи машинного обучения
2. Меры близости и расстояния.
3. Способы формирования обучающих и экзаменационных выборок. Переобученность модели.
4. Показатели качества при формировании выборок.
5. Методы оценки качества классификации.
6. Наивный байесовский метод для классификации текстовых документов.
7. Метод опорных векторов.
8. Метод центроидов для классификации текстовых документов.
9. Метод к-ближайших соседей и его модификации для классификации текстовых данных.
10. Метод деревьев решений и Random Forest для классификации данных.
11. Логистическая регрессия для решения задач классификации
12. Регуляризация в машинном обучении
13. Сравнительный анализ различных методов классификации данных.
14. Особенности обработки и анализа текстовых документов на естественном языке.
15. Модели представления текстовых документов. Векторная модель, матрица «документ-термин»
16. Методы выявления информативных признаков (терминов).
17. Формулы взвешивания терминов в текстовых документах.
18. Критерий взаимной информации для выявления информативных терминов
19. Критерий  для выявления информативных терминов
20. Эвристические подходы для выявления информативных признаков (терминов).
21. Сравнительный анализ методов выявления информативных признаков.
22. Предварительная обработка документальных массивов.
23. Профильные методы классификации. Алгоритм работы
24. Постановка задачи синтеза коллективов решающих правил (КРП) для увеличения точности классификации документальной информации.
25. Операция «отказ от классификации» в КРП.
26. Кластеризация данных. Постановка задачи, цели, примеры.
27. Показатели качества кластеризации
28. Иерархические алгоритмы кластеризации, их основные свойства
29. ЕМ-алгоритм кластеризации
30. Алгоритмы кластеризации K-средних и FOREL
31. Задача выявления полных и нечетких дубликатов.
32. Метод шинглов для выявления нечетких дубликатов.
33. Метод Winnowing для выявления нечетких дубликатов.
34. Коэффициент Жаро и Жаро-Винклера для выявления нечетких дубликатов.
35. Искусственные нейронные сети прямого распространения, виды функции активации.
36. Алгоритм обратного распространения ошибки в ИНС
37. Операция свертки. Сверточные нейронные сети.
38. Рекуррентные нейронные сети, их особенности и обучение
39. Особенности LSTM-сетей.