**Программа курса ИИС-2017**

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