**Расчетное задание по курсу ИИС**

Данное расчетное задание выполняется по вариантам, соответствующим номеру студента по журналу, и должно представлять собой завершенное исследование, содержащее сравнения, выводы, обоснования выбора тех или иных параметров и настроек. Для защиты задания необходимо представить распечатанный отчет и файл \*.ipynb с кодом.

В отчет весь код вставлять не нужно, достаточно показать результаты его выполнения, например в виде скриншотов.

**Варианты заданий:**

1. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы: КБС, Random forest, SVM, Логистическая регрессия, Desicion tree. Используя данные методы, сформировать КРП, в состав которого входят 3-5 классификаторов, который бы максимизировал точность при значении полноты не менее 80%.
2. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Реализовать классификатор «РО-профиль». Классификатор должен иметь настраиваемый параметр L - длина профиля. Вывести топ-15 слов каждого профиля. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных и построить модель, которая на тестовой выборке отвечает критерию: максимум точности при значении полноты не менее 75%.
3. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы НБ, Логистическая регрессия, Random forest. Используя данные классификаторы, определить, как влияет на точность добавление сносок и/или цитирований предыдущих сообщений
4. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Реализовать классификатор «НМИ-профиль». Классификатор должен иметь настраиваемый параметр L - длина профиля. Вывести топ-15 слов каждого профиля. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных и построить модель, которая на тестовой выборке отвечает критерию: максимум полноты при значении точности не менее 85%
5. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы SVM, Логистическая регрессия, КБС. Используя данные классификаторы, определить, как влияет на точность добавление заголовков сообщений и/или цитирований предыдущих сообщений
6. На выборке «fetch\_20newsgroups» произвольно выбрать 3 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы КБС, Логистическая регрессия, Random forest. Исследовать, как изменится точность классификации при добавлении еще 2 классов, еще 4 классов, при классификации всех 20 классов.
7. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Реализовать классификатор «С-С-профиль». Классификатор должен иметь настраиваемый параметр L - длина профиля. Вывести топ-15 слов каждого профиля. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных и построить модель, которая на тестовой выборке максимизирует критерий Accuracy.
8. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы: КБС, НБ, SVM, Логистическая регрессия, Полиномиальный Наивный Байесовский метод. Используя данные методы, сформировать КРП, в состав которого входят 3-5 классификаторов, который бы максимизировал критерий Accuracy.
9. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Реализовать классификатор «J-профиль». Классификатор должен иметь настраиваемый параметр L - длина профиля. Вывести топ-15 слов каждого профиля. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных и построить модель, которая на тестовой выборке отвечает критерию: максимум точности при значении полноты не менее 85%
10. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы: КБС, Random forest, SVM, Логистическая регрессия, НБ. Используя данные методы, сформировать КРП, в состав которого входят 3-5 классификаторов, который бы максимизировал точность при значении полноты не менее 80%.
11. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы КБС, Random forest, SVM. Используя данные классификаторы, определить, как влияет на точность добавление заголовков сообщений и/или сносок
12. На выборке «fetch\_20newsgroups» произвольно выбрать 4 класса, состоящие только из основного текста сообщения. Используя все знания об особенностях текстовых документов, осуществить предварительную обработку данных. Настроить классификаторы КБС, Логистическая регрессия, Random forest. Исследовать, как изменится точность классификации при использовании только заголовков сообщений, только заголовков и сносок, только заголовков и цитирований предыдущих сообщений