**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**“Национальный исследовательский университет ИТМО”**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

­­

**Использование машинного обучения для выбора оптимальных параметров модели СЖИГАНИЯ ГАЗОВОГО ТОПЛИВА В ПАРОГЕНЕРАТОРЕ**

Автор\_\_\_Дятлов Михаил Александрович\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, Имя, Отчество) (Подпись)

Направление подготовки (специальность)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(код, наименование)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Квалификация \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(бакалавр, магистр, инженер)\*

Руководитель ВКР\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, И., О., ученое звание, степень) (Подпись)

Санкт-Петербург, 2020 г.

Обучающийся\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Дятлов Михаил Александрович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО полностью)

Группа\_\_\_\_\_К3420\_\_\_\_\_Факультет/институт/кластер\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Направленность (профиль), специализация \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Консультант (ы):

а) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, И., О., ученое звание, степень) (Подпись)

б) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, И., О., ученое звание, степень) (Подпись)

ВКР принята “\_\_\_\_”\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

Оригинальность ВКР \_\_\_\_\_\_\_\_\_\_\_\_\_\_%

ВКР выполнена с оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата защиты “\_\_\_\_”\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

Секретарь ГЭК \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО) (подпись)

Листов хранения \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Демонстрационных материалов/Чертежей хранения \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО"**

**АнНотация**

**ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ**

**Обучающийся**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Дятлов Михаил Александрович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО)

**Наименование темы ВКР:**\_\_\_\_\_\_\_\_ использование машинного обучения для выбора\_\_\_\_\_\_

оптимальных параметров модели сжигания газового топлива в парогенераторе\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Наименование организации, где выполнена ВКР**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ХАРАКТЕРИСТИКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ**

1 Цель исследования\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2 Задачи, решаемые в ВКР \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3 Число источников, использованных при составлении обзора\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4 Полное число источников, использованных в работе \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5 В том числе источников по годам

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Отечественных** | | | | **Иностранных** | | |
| Последние 5 лет | От  5 до 10 лет | Более  10 лет | Последние  5 лет | | От  5 до 10 лет | Более  10 лет |
|  |  |  |  | |  |  |

6 Использование информационных ресурсов Internet\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Да, нет, число ссылок в списке литературы)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

7 Использование современных пакетов компьютерных программ и технологий (Указать, какие именно, и в каком разделе работы)

|  |  |
| --- | --- |
| **Пакеты компьютерных программ и технологий** | **Раздел работы** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

8 Краткая характеристика полученных результатов \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

9 Полученные гранты, при выполнении работы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Название гранта)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10 Наличие публикаций и выступлений на конференциях по теме выпускной работы\_\_\_\_\_\_\_

(Да, нет)

а) 1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Библиографическое описание публикаций)

2\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

б) 1\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Библиографическое описание выступлений на конференциях)

2\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Обучающийся\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО) (подпись)

Руководитель ВКР\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО) (подпись)

“\_\_\_\_\_\_\_\_\_\_”\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20\_\_\_г.

**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО"**

**УТВЕРЖДАЮ**

Руководитель ОП

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_

(Фамилия, И.О.) (подпись)

«\_\_\_\_» «\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_» 20\_\_\_\_ г.

**ЗАДАНИЕ**

**НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ**

**Обучающийся**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО полностью)

**Группа**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**Факультет/институт/кластер**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Квалификация** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(магистр, инженер, бакалавр)\*\*

**Направление подготовки** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(код, название направления подготовки)

**Направленность (профиль) образовательной программы** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Специализация** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Тема ВКР**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Руководитель**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО полностью, место работы, должность, ученая степень, ученое звание)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2 Срок сдачи студентом законченной работы до «**\_\_\_\_\_\_» «\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_» 20\_\_\_\_\_г.

**3 Техническое задание и исходные данные к работе** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**4 Содержание выпускной квалификационной работы (перечень подлежащих разработке вопросов)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**5 Перечень графического материала (с указанием обязательного материала**) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**6 Исходные материалы и пособия** *указанная литература должна быть не старше 10 лет*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**7 Дата выдачи задания «**\_\_\_\_» «\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_» 20\_\_\_\_г.

Руководитель ВКР\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

Задание принял к исполнению\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ «\_\_\_\_» «\_\_\_\_\_\_\_\_\_\_\_\_\_\_» 20\_\_\_\_г.

(подпись)

**СОДЕРЖАНИЕ**

[Введение 8](#_Toc39807021)

[1 Анализ предметной области 9](#_Toc39807022)

[1.1 Анализ данных 9](#_Toc39807023)

[1.2 Обработка первоначальных данных 10](#_Toc39807024)

[1.3 Алгоритмы машинного обучения 10](#_Toc39807025)

[1.4 Композиции алгоритмов 13](#_Toc39807026)

[1.5 Описание объекта 14](#_Toc39807027)

[2 Анализ предстоящей работы 16](#_Toc39807028)

[2.1 Задача 16](#_Toc39807029)

[2.2 Представление данных 16](#_Toc39807030)

[2.3 Упрощение функции 17](#_Toc39807031)

[2.4 Линейная регрессия 17](#_Toc39807032)

[2.5 Градиентный бустинг 18](#_Toc39807033)

[2.6 Вывод 18](#_Toc39807034)

[3 Основная часть 19](#_Toc39807035)

[3.1 Работа с реальными данными 19](#_Toc39807036)

[3.1.1 Задача 19](#_Toc39807037)

[3.1.2 Представление данных 19](#_Toc39807038)

[3.1.3 Сравнение алгоритмов 20](#_Toc39807039)

[Список литературы 22](#_Toc39807040)

Введение

В России 70% электроэнергии производится на тепловых электростанциях и котельных. В связи с этим очень важно использовать органическое топливо в ТЭС и котельных максимально эффективно. Кроме задачи уменьшения потребляемого топлива для экономической эффективности есть не менее важная задача уменьшения вредных выбросов в атмосферу, так как экологические задачи сейчас наиболее актуальны. Энерго-экологическая оптимизации процесса сжигания газового топлива в парогенераторе производится в том числе путем внедрения информационных технологий автоматического управления качеством сжигания топлива.

Научно производственная фирма «Уран-СПБ» уже 25 лет решает задачи энерго-экологической оптимизации парогенераторов. На данный момент «Факел-2010» новейшая система данной фирмы. Научными сотрудниками уже достаточно изучены технические аспекты оптимизации [1]. Так же в 2019 была написана бакалаврская работа на тему оптимизации технологического процесса парогенератора и в ней были применены методы линейного и нелинейного программирования.

Цель этой работы в изучении возможности внедрения машинного обучения в процесс энерго-экологической оптимизации сжигания топлива. Машинное обучения используется для предсказания параметров котла, что может позволить заранее применять регуляционные механизмы.

# Анализ предметной области

## Анализ данных

Анализ данных – это совокупность методов и алгоритмов из области математики и информатики для обработки экспериментальных данных и представления их в оптимальном виде для извлечения ценной информации и принятия решений.

Алгоритмы машинного обучения на данный момент плохо работают с просто пластами произвольных данных, поэтому перед построением алгоритма, решающего задачу, нужно обработать данные и представить их в оптимальном виде. Такая предобработка называется препроцессингом (рис. 1).

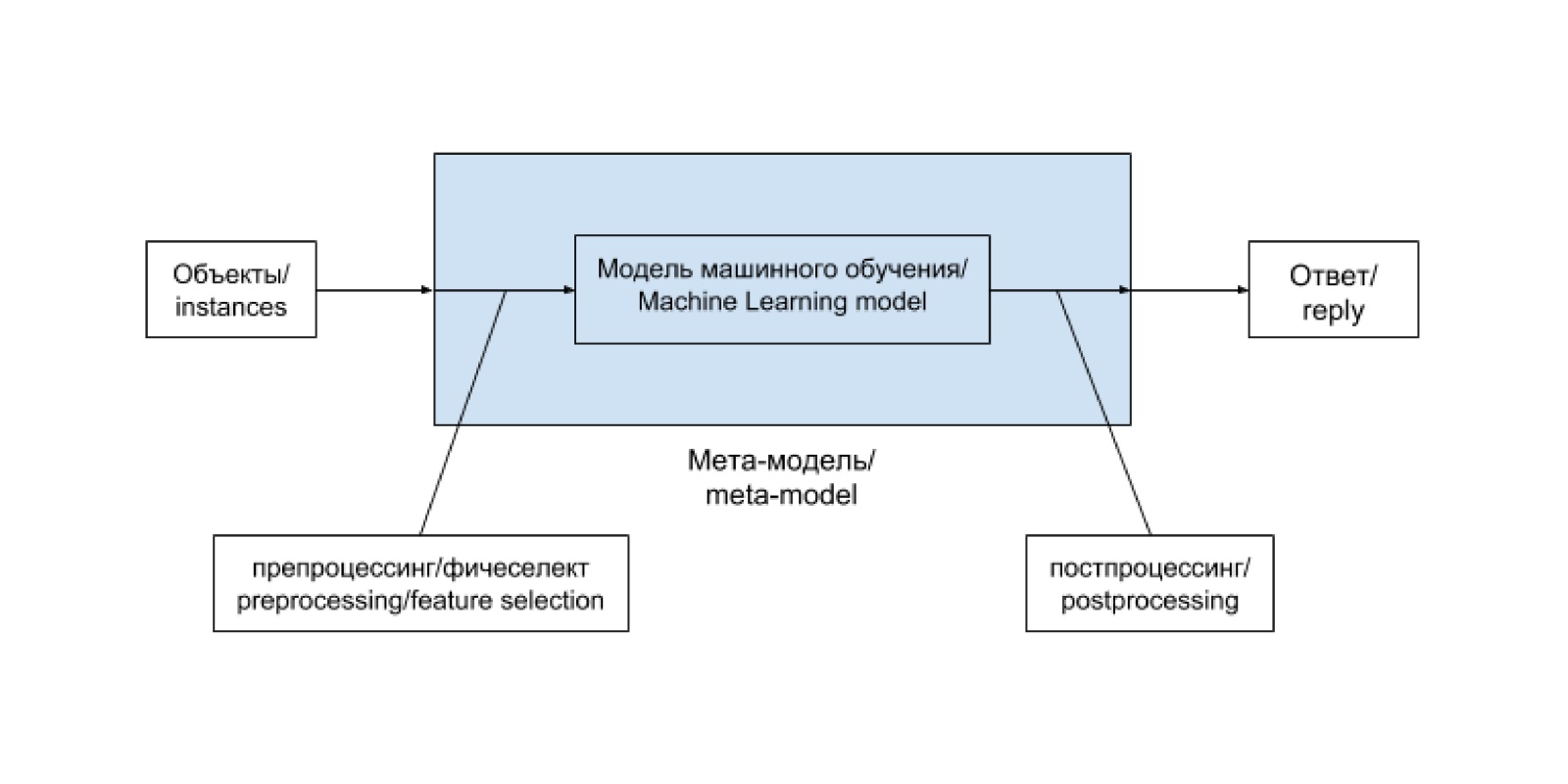


Схема решения задачи методами анализа данных.

На этапе препроцессинга выборка разбивается на обучающую, по которой модель обучается, и контрольную, на которой проверяется правильность предсказаний модели. Далее от изначальной выборке чаще всего бралось 75% на обучающую и 25% на контрольную, если не указано обратного.

## Обработка первоначальных данных

В ходе обучения модели могут возникнуть такие распространенные проблемы как: мультиколлинеарность и переобучение.

Мультиколлинеарность – это линейная зависимость между входными параметрами модели. При этом модель становится не надежной.

Переобучение – это явление, когда модель подстраивается под значения обучающей выборки, но не действенна для других данных. В этом случае модель почти невозможно применять на новых данных, что лишает ее какого-либо смысла.

Эти проблемы можно решать методами регуляризации или методом главных компонент.

L1-регуляризация – обнуляет веса неинформативных признаков.

L2-регуляризация – сокращает веса линейно зависимых признаков, но обычно не обнуляет их.

Метод главных компонент (МГК) – оставляет наиболее значимые признаки.

## Алгоритмы машинного обучения

Машинное обучение – это методы решения задачи путем обучения на большом множестве схожих задач. Оно применяется в большом количестве прикладных задач в различных областях, где есть большие объемы собранных данных и требуется решать задачи автоматизации принятия решений, предсказаний и так далее.

Машинное обучение – это недетерминированный алгоритм, от обученной модели мы ожидаем верный ответ, но не можем его требовать. Для алгоритмов машинного обучения результат не предопределен, и любая модель иногда ошибается на каких-то специфичных вариантах входных данных.

Виды моделей машинного обучения:

* + - обучение с учителем
      * классификация
      * регрессия
      * ранжирование
    - обучение без учителя
* кластеризация
* уменьшение размерности
* обучение с частичным привлечением учителя
* обучение с подкреплением

В данной работе решается задача регрессии.

### Регрессионный анализ

Регрессия – зависимость отражающая связь одной переменной от независимых переменных.

Регрессионная модель – это функция, которая приближена к исходной неизвестной функции зависимостей и построена методами регрессионного анализа.

Регрессионный анализ – решает задачу предсказания количественной целевой переменной на основе векторов признаков (других переменных). Если есть *N* строк данных, состоящих из признаков, среди которых выделяется один целевой, регрессионный анализ предсказывает какое значение целевой переменной будет в *N*+1-ой строке, если известны остальные признаки *N*+1-ой строки [2].

Приближение регрессионной модели к исходной функции осуществляется путем изменения параметров регрессионной модели и нахождения наиболее вероятных значений параметров. Регрессионная модель считается построенной, когда значения её параметров зафиксированы.

Регрессия делится на линейную и нелинейную. При линейной регрессии гипотеза происхождения данных предполагает линейную зависимость целевой переменной от признаков. Соответственной модель линейной регрессии можно представить в виде скалярного произведения параметров модели на признаки. При решении задачи методом нелинейной регрессии предполагается, что зависимость от признаков не линейна.

### Линейная регрессия

Линейная регрессия – самый распространенный алгоритм регрессионного анализа в машинном обучении. Прим­­­еняется при линейной зависимости целевого параметра от признаков. Линейная регрессия парная или простейшая, если входной параметр один и многомерная, если таких параметров множество.

Пусть у нас имеется *n* числовых признаков в выборке и­­­з *m* объектов *f*1(*x*),…,f*n*(x). По этой обучающей выборке нужно построить модель многомерной линейной регрессии. Результат – это взвешенная сумма значений признаков (весовые коэффициенты). Формула получившийся функции:

Матричное представление:

F*m×n* – матрица объекты-признаки,

*ym×1* – вектор ответов,

α*m×1* – вектор коэффициентов.

При обучении модели нужно корректировать модель с каждой ошибкой. Для уменьшения ошибок модели чаще всего используется метод наименьших квадратов, т.к. в нем происходит достаточно сильное наказание за ошибку.

Задача минимизации функционала ошибки методом наименьших квадратов для линейной регрессии:

Задача линейной регрессии состоит в том, чтобы найти вектор коэффициентов α.

### Метрики качества регрессии

Для определения того, насколько хорошо работает модель нужна метрика качества. Метрика качества предсказания алгоритма – это оценка качества предсказания в сравнении с реальными ответами. При решении каждой задачи может применяться различная метрика качества, в зависимости от того, какая особенность или характеристика предсказываемого объекта или предсказания наиболее существенна в рамках самой задачи и сферы ее применения. Основные симметричные метрик качества для регрессии:

* средняя абсолютная ошибка,
* средняя процентная ошибка,
* среднеквадратичная ошибка,
* корень из среднеквадратичной ошибки,
* коэффициент детерминации.

Средняя абсолютная ошибка – эта метрика измеряет среднее значение отклонения предсказания от реального значения по абсолютной величине, то есть непосредственно по величине параметра. Средняя абсолютная ошибка вычисляется по формуле:

Где α(*xi*) – значение, предсказанное моделью на *i*-ом объекте, *yi* – реальное значение целевого параметра объекта выборки.

Средняя процентная ошибка – показывает величину отклонения относительно действительного значения в процентах.

Формула средней процентной ошибки

Среднеквадратичная ошибка – вычисляет средний квадрат отклонения предсказания от *y*. Преимущество над средней абсолютной ошибкой в том, что ошибки гораздо сильнее влияют на показатель ошибки, это полезно, если важна точность модели.

Формула среднеквадратичной ошибки:

Корень из среднеквадратичной ошибки – вычисляется среднеквадратичная ошибка, а потом из берется корень получившегося значения. Так влияние ошибки все еще существенно, но единица измерения остается такой же как у ответов.

Формула корня из среднеквадратичной ошибки:

Коэффициент детерминации – это доля разброса значений целевой переменной.

Формула коэффициента детерминации:

ȳ - среднеарифметическое значение целевого параметра по всем объектам.

Коэффициент детерминации может принимать значения:

* 0 ≤ R2 ≤ 1 – для разумных моделей
* R2 = 1 – идеальная модель
* R2 = 0 – модель на уровне константной (константная модель – выдает всегда среднее значение целевой переменной в обучающей выборке)
* R2 < 0 – модель хуже константной

Приемлемой для линейной регрессии с константной считается модель с коэффициентом дисперсии больше 0,5. При коэффициенте детерминации больше 0,8 качество модели можно считать хорошим. Значение в единицу означает, что модель в точности описывает исходную зависимость.

## Композиции алгоритмов

При разбиении выборки для множества алгоритмов, для улучшения качество композиции применяются такие методы:

* Разбиение выборки на случайные подвыборки.
* Случайное изменение весов объектов.
* Выделение случайных подмножеств признаков для базовых алгоритмов.
* Применение различных видов базовых алгоритмов.

Бэггинг – обучает базовые алгоритмы на подвыборках, длина которых равна длине основной выборки, но состоит она из неполного множества объектов первоначальной выборки, в которой существуют повторения объектов [3, 4].

Метод случайных подпространств – базовые алгоритмы обучаются по случайным подмножествам признаков, число признаков определяется заранее и одинаково для каждого алгоритма.

Метод случайных подпространств и бэггинг можно объединить в один метод – случайный лес.

Случайный лес – бэггинг над решающими деревьями. Решающее дерево – дерево из вопросов, в котором на вершинах расположены условия а на листах прогнозы.

Простое голосование – композиция классификаторов, при которой объект относится к тому классу, который был выбран большинством классификаторов. Если вероятность правильного определения у каждого классификатора больше 0,5, то точность композиции возрастает с увеличением числа базовых классификаторов в ней.

Взвешенное голосование – отличается от простого голосования тем, что каждый базовый алгоритм имеет собственный весовой коэффициент.

Бустинг – способ построения линейной композиции базовых алгоритмов, при котором каждый следующий базовый алгоритм настраивается так, что бы покрывать слабые стороны предыдущих алгоритмов.

И бэггинг и бустинг строят композиции из уже известных базовых алгоритмов, но в отличии от бэггинга бустинг:

1. Позволяет применять произвольную функцию потерь, следовательно может быть применен не только для задач классификации, но и для задач регрессии.
2. Использует взвешенное голосования.

Градиентный бустинг – наиболее популярный алгоритм, включающий в себя функционал множества предшествующих алгоритмов бустинга.

Функция потерь – это функция характеризующая потери при не верном ответе, является мерой расхождения между правильным ответом и предсказанием.

Пусть существует функция потерь *L*(*b*, *y*), где *b* – оценка, которую дает композиция, а *y* – правильный ответ. Тогда задача минимизации функционала ошибки для градиентного бустинга выглядет как

Где T – количество базовых алгоритмов в композиции.

При решении задачи минимизации функционала ошибки алгоритм идет последовательно по каждому базовому алгоритму. При каждом новом шаге алгоритм градиентного бустинга подбирает новый базовый алгоритм и коэффициент так, чтобы приблизиться к антиградиенту функции потерь.

## Программная реализация

Используемый язык программирования для реализации: *Python*. Функции работа с данными и реализованные алгоритмы машинного обучения взяты из сторонних библиотек: *NumPy*, *scikit-learn* и *pandas*.

*NumPy* – библиотека фундаментальных научных вычислений, это мощный инструмент для математических операций в языке *python*. Одним из важнейших для дальнейшей работы инструментов является качественные *N*-мерные массивы и простая работа с ними. Так как данные удобнее всего представлять в матричном виде, *NumPy* важное средство для работы в сфере анализа данных в языке программирования python.

*Pandas* – это библиотека, которая использую функционал библиотеки *NumPy* упрощает работу с массивами данных. Благодаря библиотеке pandas можно легко преобразовать данные из множества форматов, таких как: «csv», «xls» и т.д., в матрицы из библиотеки *NumPy*.

*Scikit-learn* – основная библиотека для работы с машинным обучением в языке python. Библиотека с открытым исходным кодом. *Scikit-learn* содержит огромное количество реализованных методов машинного обучения. От классов различных видов моделей регрессии, классификации и кластеризации до методов их оценивания и методов пре- и постпроцессинга.

Основные методы и классы, использующиеся в работе:

*sklearn.linear\_model.LinearRegression* – Класс реализованной линейной регрессии, обучающуюся с применением метода наименьших квадратов. Как и все алгоритмы машинного обучения в библиотеке *scikit-learn* обучается вызовом метода *fit* с переданными в параметры матрицей признаков *X* и матрицей ответов *y*. Также имеет метод *score*, который принимает матрицы признаков и ответов тестовой выборки и возвращает коэффициент детерминации.

*sklearn.linear\_model.ElasticNet* – Такой же класс линейной регрессии, но внутри себя уже имеет L1 и L2 регулизации.

*sklearn.linear\_model.ElasticNetCV* – Представляет собой такой же класс, как «ElasticNet», но с добавление метода кросс-валидации.

*sklearn.ensemble.GradientBoostingRegressor* – Класс, реализующий метод градиентного бустинга над линейной регрессией. В параметрах можно задать (*n\_estimators*) количество шагов бустинга, выбрать функцию потерь и задать другие свойства.

*sklearn.model\_selection.train\_test\_split* – для простого разбиения данных на тестовую и обучающую выборки. В параметрах метода можно задать размер тестовой выборки, также можно выбрать перемешивать данные между собой или нет.

*pandas.read\_excel* – для чтения данных из формата «xls» и объединения столбцов «Дата» и «Время» в один через параметр «parse\_dates».

*pandas.concat* – для объединения датасетов в один.

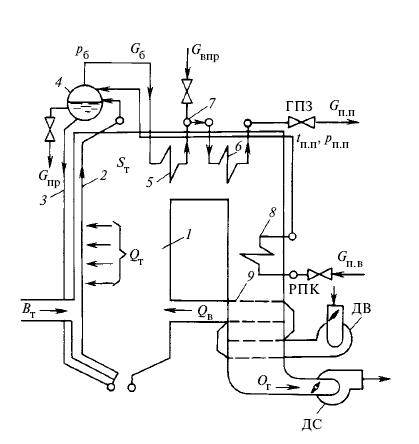
## Описание объекта

Котельная – технологический комплекс, предназначенный для выработки пара и горячей воды. Осуществляется это за счет нагрева воды в котле котельной. Работать котельные могут для обеспечения производственных нужд или для отопления. Для нагрева могут использоваться различные виды топлива: газ, жидкое топливо или твердое топливо. Котельные используются при централизованном тепло- и пароснабжении или при местном теплоснабжении зданий.

Основным устройством в котельной является котел, котлы бывают:

* Паровые
* Водогрейные
* Смешанные
* Диатермические

Котлоагрегат (котел) – обеспечивает нагрев непрерывно поступающей воды путем сжигания органического топлива.



Принципиальная технологическая схема барабанного котла.

ГПЗ – главная паровая задвижка; РПК – регулирующий питательный клапан; 1 – топка; 2 – циркуляционный контур; 3 – опускные грубы; 4 – барабан; 5, 6 – пароперегреватели; 7 – пароохладитель; 8 – экономайзер; 9 – воздухоподогреватель.

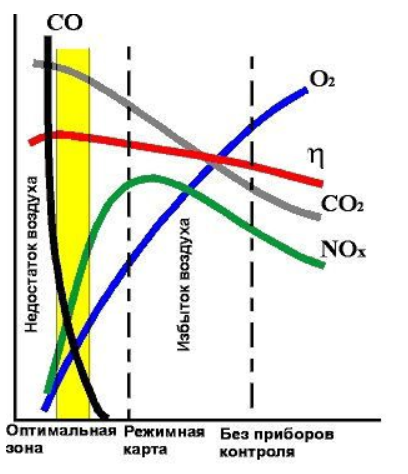
Сжигание топлива производится в топке. Жар от топлива и образующегося дыма нагревает воду.

## Оптимизируемый процесс

Основная задача – энерго-экологическая оптимизация процесса работы котлоагрегата.

Энергетическая оптимизация – по сути задача оптимизации кпд. Основной ресурс котельной – это топливо. При увеличении кпд, меньше топлива расходуется в пустоту. Основное явление, снижающее кпд то химический недожог. Химический недожог регулируется коэффициентом избытка воздуха.

Экологическая оптимизация – уменьшение вредных выбросов котлоагрегата.

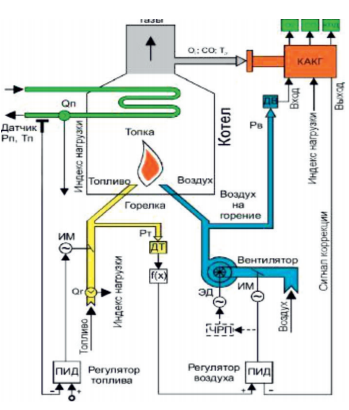


Зависимости выбросов и КПД от коэффициента избытка воздуха.

## Автоматизация регулирования

Для обеспечения лучшей работы котлоагрегата не малую роль играет автоматизация. Большинство котельных работают по режимным картам, которые не учитывают множества динамических факторов. При этом множество котельных завышают количество подаваемого воздуха, для предотвращения химического недожога, чем увеличивают количество вредных выбросов. При этом есть возможность, не теряя энергоэффективность снизить количество воздуха и уменьшить выбросы. Также котлоагрегаты часто работают на ручном регулировании, работники определяют подачу воздуха на глаз, по виду пламени котла. Такое устройство работы приводит к перерасходу топлива и повышенному выбросу загрязняющих веществ в атмосферу.

Контроль подачи воздуха и топлива можно осуществлять автоматизировано. Осуществляется за счет контроля положений исполнительных механизмов подачи топлива и газа и частоты оборотом вентиляторов.



Принципиальная схема одной из автоматизированных систем управления качеством сжигания топлива в котлоагрегате.

Благодаря автоматизации можно:

* Экономить от 6% до 12% топлива.
* Снизить выбросы оксидов азота в атмосферу на 30-40%.
* Снизить энергопотребление вентиляторов на 25-50%.
* Повысить надежность.

# Основная часть

## Работа с реальными данными

### Задача

Работа с реальными данными с датчиков котельной и построение модели регрессии для расхода газа и расхода воды (возможно концентрации СО и О2).

### Представление данных

Реальные данные представляют из себя таблицы замеров с датчиков. Данные полные (нет пропущенных ячеек). Данные взяты за определенный промежуток, в котором каждую минуту собирались данные сс датчиков и записывались в таблицу. Датчики и приборы замеряют более тридцати параметров (таблица 2.1).

Таблица 2.1 Параметры и их обозначения

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Описание параметра | Кр  Обозн | Х  мин | Х  минР | Х  максР | Х  макс |
| Температура воды на входе, гр. Ц | Твв | 0 | 20 | 150 | 180 |
| Температура воды после экономайзера, гр.Ц | Твэ | 0 | 20 | 150 | 180 |
| Давление воды после экономайзера, кг/см2 | Рвэ | 0 | 2 | 5 | 10 |
| Расход воды, м3/ч | Fв | 0 | 2 | 8 | 10 |
| Положение исполнительного механизма воды, % | ИМП | 0 | 0 | 100 | 100 |
| Уровень воды в барабане котла, мм | Lб | 0 | 25 | 75 | 100 |
| Давление пара в барабане, кгс/см2 | Рпб | 0 | 0 | 13 | 16 |
| Расход пара, т/ч | Fп | 0 | 5 | 14 | 15 |
| Температура воздуха на горение, гр.Ц | Тнв | –50 | –40 | 40 | 50 |
| Частота электропитания вентилятора, Гц | fвент | 0 | 0 | 50 | 60 |
| Положение исполнительного механизма воздуха, % | ИМВ | 0 | 0 | 100 | 100 |
| Давление воздуха общее, кПа | Рво | 0 | 5 | 200 | 250 |
| Давление воздуха в горелке 1 (левой), кПа | Рвл | 0 | 5 | 200 | 250 |
| Давление воздуха в горелке 2 (правой), кПа | Рвп | 0 | 5 | 200 | 250 |
| Расход газа, м3/час | Fг | 0 | 300 | 1300 | 2000 |
| Температура газа, гр.Ц | Тг | –20 | –15 | 30 | 40 |
| Положение исполнительного механизма газа, % | ИМГ | 0 | 0 | 100 | 100 |

Сырые данные представлены в виде таблицы в формате программного обеспечения *Microsoft Excel*.

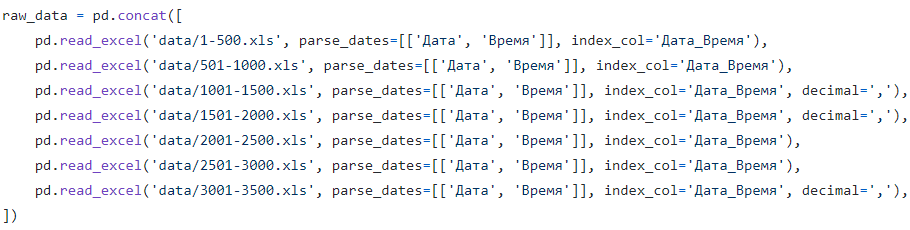
Анализ предстоящего преобразования данных:

1. Данные нужно преобразовать из файлов формата «.xls» к матричному виду.
2. Данные лежат в разных файлах, требуется объединить в один датасет.
3. Нужно выбелить какие столбцы будут целевыми, а какие представляют собой параметры для предсказания.
4. Для каждого выбранного целевого параметра убрать столбцы с прямой корреляцией (применить методы регуляризации).
5. Выборка содержит дату в одном столбце и время в другом, нужно объединить и обеспечить корректную работу с информацией о дате и времени замеров.
6. В некоторых файлах вместо точки как разделителя используется запятая.

В качестве целевых переменных выступают: содержание оксида углерода (CO), содержание кислорода после котла (O2к), расход воды (Fв) и расход газа (Fг). Признаки, выбранные для первого шага: положение исполнительного механизма воды (ИМП), положение исполнительного механизма газа (ИМГ), частота вентилятора, падающего воздух (fвент) и частота вентилятора дымососа (fдым).

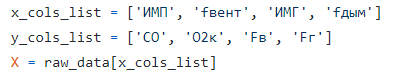
### Программная реализация

Для чтения из файла «.xls» используется метод *read\_excel* а для объединения в один датасет метод *concat*, оба метода из библиотеки pandas. Задать разделитель и объединить дату и время можно с помощью параметров *decimal* и *parse\_dates* метода *read\_excel*.



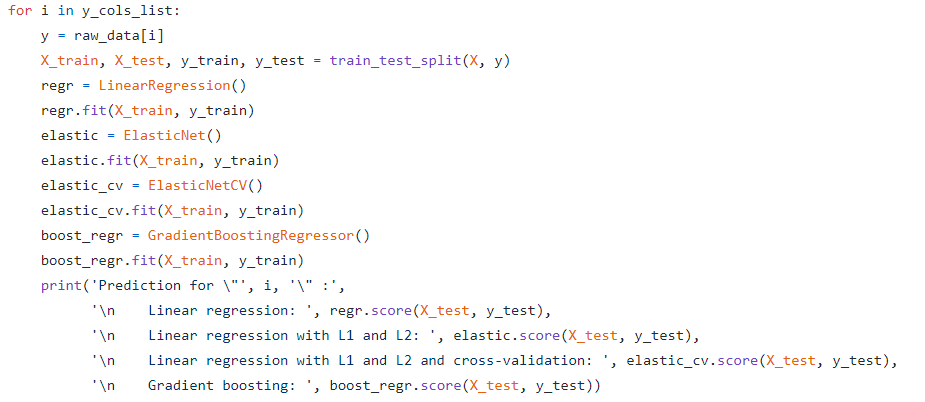
Программная реализация 1.1.

Целевые переменные и признаки разделяются очень просто. В данном случае и для матрицы признаков, и для матрицы ответов были созданы массивы строк с названиями колонок нужных параметров. Сразу строится матрица признаков. Делается это так же, как в массиве происходит обращение по индексу, но вместо индекса передается массив с названиями столбцов. Так из матрицы *raw\_data* библиотека *NumPy* возвращает матрицу признаков.



Программная реализация 1.2.

Оставшаяся часть кода – это цикл по элементам массива *y\_cols\_list*. В этом цикле для каждого элемента массива создается новая матрица ответов из одного столбца. Далее, с помощью уже описанного метода *train\_test\_split*, обе матрицы разбиваются на обучающую и тестовую. Потом объявляются объекты описанных в первой части классов алгоритмов машинного обучения. Каждый объект обучается на обучающей выборке. И на конец в функции print вызываются значения коэффициентов детерминации для переданных в параметры метода score тестовых матриц. Метод *score* сразу предсказывает значения для переданной матрицы признаков, сравнивает их с матрицей ответов и считает и возвращает коэффициент детерминации. Таким образом для проверки качества модели даже не нужно отдельно вызывать метод предсказания. Далее, когда потребуется предсказывать параметры для новых данных, конечно, будет вызываться метод *predict*. Но на данном этапе важна только проверка качества модели.



Программная реализация 1.3.

### Сравнение алгоритмов

Для выбранных целевых параметров были обучены модели линейной регрессии, линейной регрессии с применением L1 и L2 регуляризации, линейной регрессии с применением L1 и L2 регуляризации и кросс-валидации и градиентного бустинга.

Значения коэффициентов детерминации для различных моделей при различных представлениях выборки представлены в таблице (2.2).

Таблица 2.2 – Сравнение моделей.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Коэффициент детерминации | | | |
| Целевая переменная | CO | O2к | Fв | Fг |
| **Не перемешанная выборка** |  |  |  |  |
| Линейная регрессия | 0,11 | 0,14 | 0,87 | 0,67 |
| Линейная регрессия с L1 и L2 регулизациями | 0,37 | 0,05 | 0,87 | 0,66 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,11 | -0,02 | 0,87 | 0,67 |
| Градиентный бустинг | -0,78 | -0,39 | 0,91 | 0,54 |
| **Перемешанная выборка** |  |  |  |  |
| Линейная регрессия | 0,29 | 0,09 | 0,89 | 0,69 |
| Линейная регрессия с L1 и L2 регулизациями | 0,30 | 0,18 | 0,89 | 0,68 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,3 | 0,1 | 0,89 | 0,69 |
| Градиентный бустинг | 0,6 | 0,77 | 0,94 | 0,78 |
| **Перемешанная выборка с добавлением столбца «Дата\_Время» в параметры** |  |  |  |  |
| Линейная регрессия | 0,45 | 0,26 | 0,9 | 0,72 |
| Линейная регрессия с L1 и L2 регулизациями | 0,42 | 0,23 | 0,9 | 0,71 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,01 | 0 | 0 | 0,12 |
| Градиентный бустинг | 0,83 | 0,96 | 0,94 | 0,76 |

­

Из таблицы (1) хорошо видно, как различная работа с данными и применение отличающихся алгоритмов влияет на предсказание содержания кислорода после котла (O2к). Для предсказания этого параметра важным оказывается, что бы данные в выборке располагались не последовательно, а были перемешаны между собой. При этом наиболее эффективным оказался градиентный бустинг.

### Вывод

Исходя из первой задачи можно видеть какую большую роль играет работа с данными в машинном обучении. Действительно, линейная регрессия известна достаточно давно, но именно наличие больших объемов данных и вычислительных мощностей сделало машинное обучение возможным в том виде, в котором оно есть сейчас. Поэтому было важно изучить поведение реальных данных, с которыми будет производиться работа.

## Улучшение модели

### Задача

Улучшение модели путем подбора новых параметров мониторинга и добавления их в матрицу признаков.

### Подбор новых признаков

Важным инструментом для выбора признаков является матрица корреляции. Матрицу корреляции можно построит средствами библиотеки pandas, вызвав у матрицы выборки метод corr. В случае с представленными данными с котельной, матрица корреляции уже была построена, поэтому для удобства использоваться будет она.

Матрица корреляции – это матрица или таблица показывающая зависимость переменных между собой. Заголовки столбцов и строк – заголовки параметров, на пересечении находится зависимость между двумя параметрами.

В связи с тем, что количество признаков и целевых переменных не слишком большое, матрица корреляции представлена в удобном виде и процесс местами требует аналитики, новые признаки подбирались вручную.

Сначала для каждого целевого параметра были выбраны параметры со значимым коэффициентом корреляции. Так как эти параметры наиболее значимы и имеют большее влияние на целевую переменную. Далее среди выбранных признаков были найдены признаки, корреляция между которыми превышает значение 0,9. Если два признака сильно коррелируют между собой, выбирался тот признак, который больше влиял на целевую переменную.

Таким образом решалась задача оптимизации модели. Данный метод строится на основном свойстве линейной регрессии, описанном выше. Признаки должны коррелировать с целевой переменной, так как считается, что она зависима от них. При это между собой признаки независимы. При зависимости между признаками возникает проблема мультиколлинеарности, а за ней и проблема переобучения.

### Выбранные признаки.

Для расхода газа (Fг): давление воды после экономайзера (Pвэ), давление пара в барабане (Рпб), положение исполнительного механизма газа (ИМГ), частота вращения вентилятора дымососа (fдым), давление в дымоходе после экономайзера (Pдэ), общее давление газа (Рго).

Для расхода воды (Fв): давление воды после экономайзера (Pвэ), давление пара в барабане (Рпб), температура воды после экономайзера (Tвэ), положение исполнительного механизма воды (ИМП), температура дымовых газов после котла (Tдк).

Для содержания оксида углерода (СО): температура газа (Tг), Разрежение в дымоходе после экономайзера (Pдэ), общее давление газа (Рго), частота вращения вентилятора дымососа (fдым), дата и время (Дата\_Время).

Для содержания кислорода после котла ('О2к'): температура газа (Tг), Разрежение в дымоходе после экономайзера (Pдэ), общее давление газа (Рго), частота вращения вентилятора дымососа (fдым), дата и время (Дата\_Время).

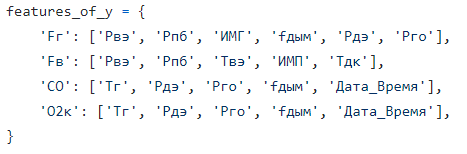
### Программная реализация.

В предыдущем подразделе было выявлено, что дата и время в признаках существенно улучшает модели предсказания содержания оксида углерода и кислорода. Поэтому они добавлены в новых признаках. Для преобразования их к числовому виду применяется метод to\_numeric из библиотеки pandas. Также ранее дата и время были обозначены как колонка индексов для данных, теперь же следует убрать это обозначение при чтении данных, индексы будут присвоены автоматически, от 0 по возрастанию.



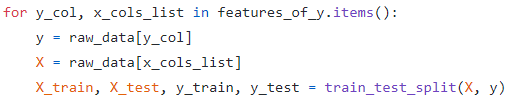
Программная реализация 2.1.

В связи с тем, что для каждой целевой переменной теперь выбраны свои признаки объявление массивов как в первом подразделе не подходит. Для нового подхода была применена структура данных языка *python* – словарь. В словаре ключом является целевая переменная, а значением массив обозначений признаков. Также объявление матрицы признаков теперь не может происходить сразу.



Программная реализация 2.2.

Далее также идет цикл, но уже не по элементам массива обозначений целевых признаков, а по парам ключ-значение словаря. Для каждого ключа создается матрица *y* а по значению матрица *X*. По этим матрицам создаются обучающие и тестовые выборки. Далее код не отличается от кода в первом подразделе.



Программная реализация 2.3.

### Новые показатели моделей

Для новых признаков были также обучены модели. Новые значения представлены в таблице 2.3.

Таблица 2.3 – Качество моделей с новыми признаками.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Коэффициент детерминации | | | |
| Целевая переменная | CO | O2к | Fв | Fг |
| **Не перемешанная выборка** |  |  |  |  |
| Линейная регрессия | 0,11 | 0,14 | 0,87 | 0,67 |
| Линейная регрессия с L1 и L2 регулизациями | 0,37 | 0,05 | 0,87 | 0,66 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,11 | -0,02 | 0,87 | 0,67 |
| Градиентный бустинг | -0,78 | -0,39 | 0,91 | 0,54 |
| **Перемешанная выборка** |  |  |  |  |
| Линейная регрессия | 0,29 | 0,09 | 0,89 | 0,69 |
| Линейная регрессия с L1 и L2 регулизациями | 0,30 | 0,18 | 0,89 | 0,68 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,3 | 0,1 | 0,89 | 0,69 |
| Градиентный бустинг | 0,6 | 0,77 | 0,94 | 0,78 |
| **Перемешанная выборка с добавлением столбца «Дата\_Время» в параметры** |  |  |  |  |
| Линейная регрессия | 0,45 | 0,26 | 0,9 | 0,72 |
| Линейная регрессия с L1 и L2 регулизациями | 0,42 | 0,23 | 0,9 | 0,71 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,01 | 0 | 0 | 0,12 |
| Градиентный бустинг | 0,83 | 0,96 | 0,94 | 0,76 |

­

### Вывод

## Минимизация расхода газа

Список литературы

1. Новиков, О. Н. Как повысить эффективность работы котельной / О.Н. Новиков, А.Н. Окатьев, Н.А. Антонова. // Промышленно–строительное обозрение – 2013 – № 8 – с.61–63.
2. Разинков Е.В. Машинное обучение // Электронная библиотека ИВМиИТ-КФУ. – 2015. – 25 апреля [Электронный ресурс]. URL: <https://vmkhelp.ru/wp-content/uploads/2017/09/machine_learning_at_kfu_25_04_15.pdf> (дата обращения: 7.05.2020).
3. Воронцов К.В. Лекции по алгоритмическим композициям // Профессиональный информационно-аналитический ресурс. – 2012. – 7 октября [Электронный ресурс]. URL: <http://www.machinelearning.ru/wiki/images/0/0d/Voron-ML-Compositions.pdf> (дата обращения: 7.05.2020).
4. Кашницкий Ю. С., Игнатов Д. И. Ансамблевый метод машинного обучения, основанный на рекомендации классификаторов // Интеллектуальные системы. Теория и приложения. – 2015. – Т. 19. – №4. – С.45 – 47.