**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).

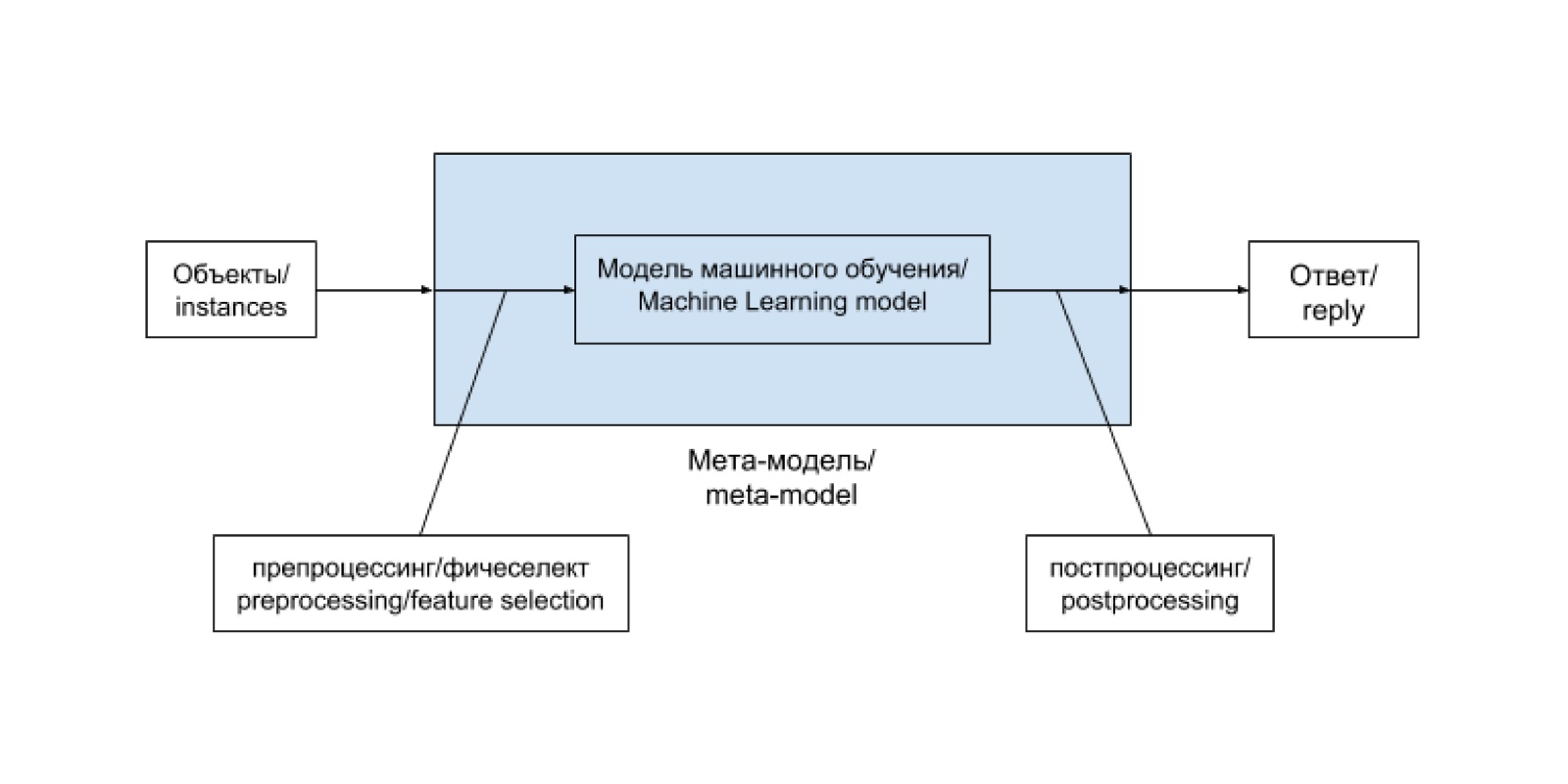


Схема решения задачи методами анализа данных.

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.

Регрессия – 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, регрессия предсказывает какое значение целевой переменной будет в *N*+1-ой строке, если известны остальные признаки *N*+1-ой строки [2].

Линейная регрессия – самый распространенный алгоритм решения задачи регрессии 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.

Для определения того, насколько хорошо работает модель нужна метрика качества. Метрика качества предсказания алгоритма – это оценка качества предсказания в сравнении с реальными ответами. Основные симметричные метрик качества для регрессии:

* среднеквадратичная ошибка,
* средняя абсолютная ошибка,
* коэффициент детерминации.

Формула коэффициента детерминации:

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

## 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.

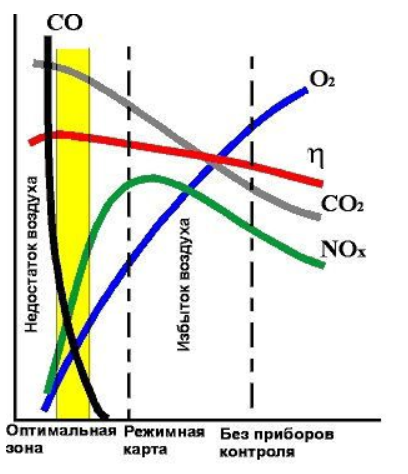
Функция потерь – это функция характеризующая потери при не верном ответе, является мерой расхождения между правильным ответом и предсказанием.

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, что бы приблизиться к антиградиенту функции потерь.

## Описание объекта



Зависимости выбросов и КПД от коэффициента избытка воздуха.

# Анализ предстоящей работы

## Задача

Изначальная задача состояла в предсказании графика процесса горения. Для

этой задачи была поставлена искусственная подзадача по предсказанию графика функции

для разработки прототипа и анализа сложностей первоначальной задачи.

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

## Представление данных

Для представления данных были использованы два варианта.

Для каждой пары параметров *a* и *b* при *x* от -100 до 100 представляется данные о значениях параметра *a*, *b*, *x* и полученном *y*. Таким образом строится определенное количество «графиков», для каждого из которых в выборке данных получается по 201 строчке, по одной для каждого значения *x* в выбранном интервале. Значения *a* и *b* выбирались в каждом «графике» случайными в интервале от -500 до 500. Также возможно для каждого значения *a* выбирать несколько разных значений *b* и наоборот. Для того что бы были представлены случаи с неизменным одним параметром, для лучшего определения влияния второго параметра.

Второй опробованный вариант – это разбиение на группы и построение модели для каждой группы. Это можно визуализировать и лучше понять таким образом: регрессионные методы машинного обучения для решения экономической задачи предсказания цены на что-либо имеют в выборке значения цены за прошедший промежуток времени, то есть не завершённый график *y* от *x*, и по этим данным предсказывают какое будет следующее значение *y*, а так как в нашем случае каждое значение *y* это уже како-то график, но завершенный, мы можем расположить их в новой плоскости *z* и тогда нам нужно предсказать какой будет следующий график в этой плоскости *z*. Но так как модель машинного обучения на выходе выдает одно значение, то для каждого *x* мы строим новый график *y* от *z* и предсказываем значение *y* от нового параметра *z*, а потом из всех предсказанных значений для каждого *x* получается требуемый график.

## Упрощение функции

В нашей задаче есть какая-то сложная функция с большим количеством параметров, каждый из который в разной степени влияет на значение *y* в точке *x*. По этой функции строится график *y* от *x,* и наша задача предположить, как будет выглядеть график при заданных параметрах.

Первый шаг — это упростить неизвестную нам функцию с множеством параметров до функции с меньшим количеством параметров и более простой зависимостью *y* от них.

## Линейная регрессия

При обучении линейной регрессии на сырых данных коэффициент детерминации модели примерно равен 0,5567. Это достаточно низкая точность. Я предположил, что проблема в том, что модель не принимает во внимание параметр *x*. Поэтому применим второй метод представления данных.

При разбиении данных, средний коэффициент детерминации моделей (для каждого *x* своя модель) равен 1.

К сожалению, линейная регрессия работает хорошо для линейных зависимостей значения от параметров. Если мы заменим формулу (1.6) на

коэффициент детерминации даже при втором способе представления данных падает до 0,5479.

## Градиентный бустинг

Для формулы (1.6) коэффициент детерминации градиентного бустинга примерно равен 0,9931, средне значение у моделей для каждого X примерно равно 0,9984.

Для формулы (1.7) коэффициент детерминации градиентного бустинга примерно равен 0,97, средне значение у моделей для каждого X примерно равно 0,88.

## Вывод

По завершению анализа были применены теоретические знания и получены практические навыки при реализации алгоритмов анализа данных для искусственного примера. Благодаря этому требуемые действия при решении задачи на реальных данных заранее более очевидны.

# Основная часть

## Работа с реальными данными

### Задача

По реальным данным нужно предсказывать какие-либо выбранные параметры из множества параметров с различными зависимостями. В том числе некоторые переменные могут быть как целевыми параметрами, так и признаками для предсказания других параметров.

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

### Представление данных

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.

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. В качестве целевых переменных было решено выбрать: разрежение в дымоходе после котла (Pдк), температуру дымовых газов после котла (Tдк) и содержание кислорода после котла (O2к). Остальные параметры, за исключением даты и временя, являются признаками для будущей модели машинного обучения. При предсказании различных параметров возможны ручные изменения списка признаков.

### 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 (3.1).

Таблица 3.1 – сравнение моделей.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Коэффициент детерминации | | |
| Целевая переменная | O2к | Pдк | Tдк |
| **Не перемешанная выборка** |  |  |  |
| Линейная регрессия | 0,34 | 0,99 | 0,977 |
| Линейная регрессия с L1 и L2 регулизациями | 0,05 | 0,99 | 0,966 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | -0,02 | 0,99 | 0,98 |
| Градиентный бустинг | 0,07 | 0,9 | 0,9 |
| **Перемешанная выборка** |  |  |  |
| Линейная регрессия | 0,84 | 0,979 | 0,99 |
| Линейная регрессия с L1 и L2 регулизациями | 0,4 | 0,99 | 0,98 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | 0,82 | 0,99 | 0,98 |
| Градиентный бустинг | 0,97 | 0,98 | 0,99 |
| **Перемешанная выборка с добавлением столбца «Дата\_Время» в параметры** |  |  |  |
| Линейная регрессия | 0,7 | 0,99 | 0,98 |
| Линейная регрессия с L1 и L2 регулизациями | 0,27 | 0,99 | 0,98 |
| Линейная регрессия с L1 и L2 регулизациями и кросс-валидацией | -0,002 | 0,11 | 0,027 |
| Градиентный бустинг | 0,96 | 0,99 | 0,99 |

­

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. А столбец даты и времени только сбивает линейную регрессию, даже с регулизацией.

### Вывод

Исходя из этих двух подзадач 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.

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.