**Описание проекта: промышленность**

Чтобы оптимизировать производственные расходы, металлургический комбинат «Стальная птица» решил уменьшить потребление электроэнергии на этапе обработки стали. Для этого комбинату нужно контролировать температуру сплава. Ваша задача — построить модель, которая будет её предсказывать.

Заказчик хочет использовать разработанную модель для имитации технологического процесса. Ниже расскажем о деталях этого процесса. Их важно знать, прежде чем генерировать новые признаки.

**Описание этапа обработки**

Сталь обрабатывают в металлическом ковше вместимостью около 100 тонн. Чтобы ковш выдерживал высокие температуры, изнутри его облицовывают огнеупорным кирпичом. Расплавленную сталь заливают в ковш и подогревают до нужной температуры графитовыми электродами. Они установлены в крышке ковша.

Из сплава выводится сера (этот процесс — десульфурация), добавлением примесей корректируется химический состав и отбираются пробы. Сталь легируют — изменяют её состав — подавая куски сплава из бункера для сыпучих материалов или проволоку через специальный трайб-аппарат (от англ. tribe — «масса»).

Перед тем как первый раз ввести легирующие добавки, измеряют температуру стали и производят её химический анализ. Потом температуру на несколько минут повышают, добавляют легирующие материалы и продувают сплав инертным газом. Затем его перемешивают и снова проводят измерения. Такой цикл повторяется до достижения целевого химического состава и оптимальной температуры плавки.

Тогда расплавленная сталь отправляется на доводку металла или поступает в машину непрерывной разливки. Оттуда готовый продукт выходит в виде заготовок-слябов (от англ. slab — «плита»).

**Описание данных**

Данные состоят из нескольких файлов, полученных из разных источников:

* data\_arc\_new.csv — данные об электродах;
* data\_bulk\_new.csv — данные о подаче сыпучих материалов (объём);
* data\_bulk\_time\_new.csv — данные о подаче сыпучих материалов (время);
* data\_gas\_new.csv — данные о продувке сплава газом;
* data\_temp\_new.csv — результаты измерения температуры;
* data\_wire\_new.csv — данные о проволочных материалах (объём);
* data\_wire\_time\_new.csv — данные о проволочных материалах (время).

**Файл data\_arc\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_arc_new.csv)

* key — номер партии;
* Начало нагрева дугой — время начала нагрева;
* Конец нагрева дугой — время окончания нагрева;
* Активная мощность — значение активной мощности;
* Реактивная мощность — значение реактивной мощности.

**Файл data\_bulk\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_bulk_new.csv)

* key — номер партии;
* Bulk 1 … Bulk 15 — объём подаваемого материала.

**Файл data\_bulk\_time\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_bulk_time_new.csv)

* key — номер партии;
* Bulk 1 … Bulk 15 — время подачи материала.

**Файл data\_gas\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_gas_new.csv)

* key — номер партии;
* Газ 1 — объём подаваемого газа.

**Файл data\_temp\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_temp_new.csv)

* key — номер партии;
* Время замера — время замера;
* Температура — значение температуры.

**Файл data\_wire\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_wire_new.csv)

* key — номер партии;
* Wire 1 … Wire 9 — объём подаваемых проволочных материалов.

**Файл data\_wire\_time\_new.csv**

[Скачать файл](https://code.s3.yandex.net/datasets/data_wire_time_new.csv)

* key — номер партии;
* Wire 1 … Wire 9 — время подачи проволочных материалов.

Во всех файлах столбец key содержит номер партии. В файлах может быть несколько строк с одинаковым значением key: они соответствуют разным итерациям обработки.

Данные также находятся в тренажёре, в папке /datasets/.

**План работы**

Позади много пройденных уроков, вы уже выпускники! Уверены, что вы умеете выделять основные этапы работы и следовать намеченному плану. Тем не менее оставим вам краткий план, который вы сможете дополнить необходимыми, на ваш взгляд, нюансами.

**Шаг 1. Загрузка данных**

Загрузите данные и выполните их первичный осмотр.

**Шаг 2. Исследовательский анализ и предобработка данных**

Выполните исследовательский анализ каждого датафрейма и при необходимости выполните предобработку. Сделайте выводы об имеющихся признаках: понадобятся ли они для обучения моделей.

**Шаг 3. Объединение данных**

Объедините выбранные вами признаки в один датафрейм по ключу.

**Шаг 4. Исследовательский анализ и предобработка данных объединённого датафрейма**

Выполните исследовательский анализ объединённого датафрейма, визуализируйте распределения признаков и при необходимости выполните предобработку. Проведите корреляционный анализ. Напоминаем, что вы можете использовать не только имеющиеся признаки, но и генерировать новые.

**Шаг 5. Подготовка данных**

Выполните подготовку данных для обучения модели. Разделите данные на две выборки, при масштабировании и кодировании учитывайте особенности данных и моделей.

**Шаг 6. Обучение моделей машинного обучения**

Обучите как минимум две модели. Хотя бы для одной из них подберите как минимум два гиперпараметра.

**Шаг 7. Выбор лучшей модели**

Выберите лучшую модель и проверьте её качество на тестовой выборке.

**Шаг 8. Общий вывод и рекомендации заказчику**

Сделайте общий вывод о проделанной работе: опишите основные этапы работы, полученные результаты и дайте рекомендации для бизнеса.