

Прогнозирование температуры звезды

Вам пришла задача от обсерватории «Небо на ладони»: придумать, как с помощью нейросети определять температуру на поверхности обнаруженных звёзд. Обычно для расчёта температуры учёные пользуются следующими методами:

- Закон смещения Вина.
- Закон Стефана-Больцмана.
- Спектральный анализ.

Каждый из них имеет плюсы и минусы. Обсерватория хочет внедрить технологии машинного обучения для предсказания температуры звёзд, надеясь, что этот метод будет наиболее точным и удобным.

В базе обсерватории есть характеристики уже изученных 240 звёзд.

Характеристики

- Относительная светимость L/L_0 — светимость звезды относительно Солнца.
- Относительный радиус R/R_0 — радиус звезды относительно радиуса Солнца.
- Абсолютная звёздная величина M_V — физическая величина, характеризующая блеск звезды.
- Звёздный цвет (white, red, blue, yellow, yellow-orange и др.) — цвет звезды, который определяют на основе спектрального анализа.
- Тип звезды.

Тип звезды	Номер, соответствующий типу
Коричневый карлик	0
Красный карлик	1
Белый карлик	2
Звёзды главной последовательности	3
Сверхгигант	4
Гипергигант	5

- Абсолютная температура T (К) — температура на поверхности звезды в Кельвинах.

В этом самостоятельном проекте вам необходимо разработать нейронную сеть, которая поможет предсказывать абсолютную температуру на поверхности звезды.

Справочная информация: Светимость Солнца (англ. *Average Luminosity of Sun*) $L_0 = 3.828 \cdot 10^{26}$ Вт Радиус Солнца (англ. *Average Radius of Sun*) $R_0 = 6.9551 \cdot 10^8$ м

Алгоритм решения задачи

Шаг 1. Загрузка исходных данных из файла

/datasets/6_class.csv

Шаг 2. Исследовательский анализ

Проведите исследовательский анализ:

- количественных данных,

- категориальных данных.

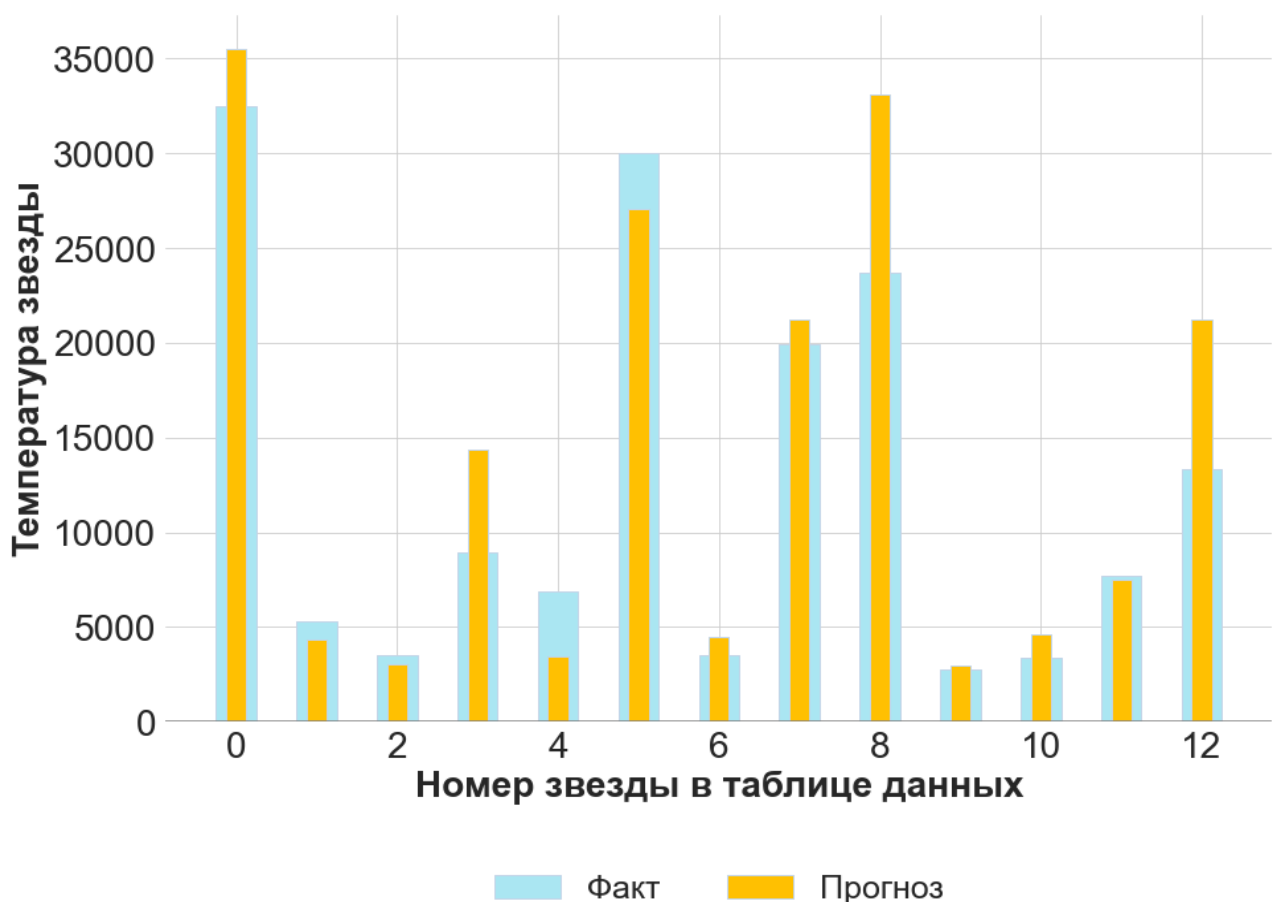
Обязательно используйте графический анализ и сделайте вывод.

Шаг 3. Подготовка данных к построению модели

1. По результату исследовательского анализа внесите корректировки, если они нужны. Сделайте вывод.
2. Если необходимо, категоризируйте исходные данные.
3. Подготовьте обучающую и тестовую выборки.
4. Проведите масштабирование количественных данных.

Шаг 4. Построение простой модели нейронной сети — baseline

1. Создайте класс для задания архитектуры нейронной сети.
2. Самостоятельно выберите количество скрытых слоёв, количество нейронов на них, функции активации на скрытых и выходном слоях. Попробуйте и сравните несколько подобных комбинаций.
3. Проведите обучение нейронной сети:
 - Создайте функцию для обучения нейронной сети.
 - Проведите обучение модели.
 - Постройте график «Факт — Прогноз», где по горизонтальной оси будут отложены условные номера звёзд, а по вертикальной — температура в Кельвинах.



- Сделайте вывод.

Шаг 5. Улучшение сети

- Создайте решение с перебором параметров нейросети. Список параметров для перебора должен включать как минимум «dropout» и «размер батча». Архитектуру нейронной сети: количество слоёв, нейронов, вид функции активации — оставьте как в *Baseline*, чтобы сравнить результат.
- Проведите обучение нейронной сети. Выведите метрику RMSE и график «Факт — прогноз». Метрика RMSE не должна превышать 4500.
- Сделайте вывод. Желательно оформить его в виде таблицы или графика.

Шаг 6. Выводы по проекту

- Кратко опишите результаты каждого типа модели.
- Напишите выводы сравнения двух моделей.