

## Задание №1

### Предыстория

Команда студентов из Новосибирского государственного университета разрабатывает алгоритмы для рекламной платформы «Сибирский клин». Анализируя данные о показах рекламы, они стремятся предсказать, кликнет ли пользователь по объявлению, чтобы оптимизировать таргетинг.

Вас отобрали в команду — поздравляем! Теперь вам предстоит обработать большой массив данных и создать модель для предсказания вероятности клика.

### Что нужно сделать

По категориальным признакам из данных о показах рекламы предсказать вероятность клика (целевой признак: 0 или 1).

### Данные

Что	Путь	Формат и описание
Обучающая выборка	data/train.csv	40 000 000 строк, столбцы: id, 22 категориальных признака (ID_01... ID_22), click (0/1)
Тестовая выборка	data/test.csv	Те же столбцы без click
Пример сабмита	data/sample_submission.csv	idx,click

### Метрика

ROC-AUC (Receiver Operating Characteristic - Area Under Curve).

#### Шкала баллов

$\text{Points} = 100 \times \max(0, \text{ROC-AUC} - 0.60) / (1 - 0.60)$

При ROC-AUC < 0.60 — 0 баллов. «Точное» решение: ROC-AUC  $\geq$  0.85.

### Система оценки

$\text{Points} = 100 \times \max(0, \text{ROC-AUC} - 0.60) / (1 - 0.60)$

При ROC-AUC < 0.60 — 0 баллов. «Точное» решение: ROC-AUC  $\geq$  0.85.

### Замечание

### Рекомендации по обработке

- **Предобработка:** Используйте one-hot encoding или hashing для категориальных признаков. Учитывайте высокую кардинальность некоторых столбцов.
- **Базовый baseline:** Логистическая регрессия или SGDClassifier (scikit-learn).
- **Оптимизация:** Попробуйте градиентный бустинг (LightGBM) или нейронные сети для повышения ROC-AUC.
- **Кросс-валидация:** Используйте 5-fold CV для оценки модели на обучающей выборке.

## Задание №2

## Предыстория

Клуб «Следопыт Закавказья» накопил архив из примерно 3000 фотографий, содержащих изображения следов лап и самих лап диких животных и птиц, собранных во время таёжных экспедиций.

Фотографии сделаны в разных условиях: часть — на снегу, часть — на грунте без снега и льда. Некоторые снимки показывают только отпечатки, другие — непосредственно лапы животных или птиц. В кадре нет изображений целых животных — только детали следов или лап на различных поверхностях.

Вам предстоит создать модель для определения вида животного по отпечатку, чтобы егеря могли своевременно отследить редких обитателей и предупредить туристов о хищниках, скрывающихся в тени сосен.

## Что нужно сделать

По фотографии отпечатка определить, какому из шести видов животных принадлежит след: медведь, птица, кошка, волк, выдра или леопард.

## Данные

Что	Путь	Формат и описание
Обучающие изображения	data/train/bear/.../fox/	JPG/PNG, произвольный размер $\leq 400 \times 400$ , имена — ID файлов
Тестовые изображения	data/test/	Те же форматы, без разметки
Карта меток	data/label_map.csv	Столбцы id, filename, Bear, Bird, Cat, Wolf, Leopard, Otter
Пример сабмита	data/sample_submission.csv	id,label (Bear - 0... Otter - 5)

## Метрика

Macro F1-score по 6 классам.

### Шкала баллов

Points =  $150 \times \max(0, F1 - 0.50) / (1 - 0.50)$

При  $F1 < 0.50$  — 0 баллов. «Точное» решение:  $F1 \geq 0.95$  (ошибка  $\leq 5\%$ ).

## Система оценки

Points =  $150 \times \max(0, F1 - 0.50) / (1 - 0.50)$

При  $F1 < 0.50$  — 0 баллов. «Точное» решение:  $F1 \geq 0.95$  (ошибка  $\leq 5\%$ ).

## Замечание

### Рекомендации по обработке

- Приведение к единой размерности: `resize` до  $224 \times 224$  + нормировка по ImageNet.
- Обязательные аугментации: случайный обрез (`RandomResizedCrop`), `flips`, `color jitter` — следы снимали в разных условиях освещения.
- Базовый baseline: fine-tune resnet18 (PyTorch), оптимизатор Adam,  $LR=1e-3$ , 20 эпох, batch size=32.

## Задание №3

## Предыстория

Ученики астрономического клуба «Орион» из Новосибирска запустили проект «Тёмные небеса». С помощью дронов «Сокол-Наблюдатель», оснащённых камерами и датчиками SQM, они совершили 1400 ночных полётов над городами Сибири, чтобы измерить уровень светового загрязнения.

Теперь необходимо автоматически определять яркость неба по фотографиям для создания общероссийской карты «Звёздный щит».

## Что нужно сделать

По изображению предсказать значение `sky_quality` (тип: `float`, единицы: `mag/arcsec²`) для тестовых снимков.

## Данные

Что	Путь	Формат и описание
Обучающие фото	<code>data/images/IMG_0001.jpg...</code>	224×224 RGB, имя = <code>IMG_####.jpg</code>
Таблица меток	<code>data/metadata.csv</code>	<code>file_name,sky_quality</code>
Тестовые фото	<code>data/test_images/</code>	Те же форматы, без меток
Пример сабмита	<code>data/sample_submission.csv</code>	<code>file_name,sky_quality</code>

## Метрика

RMSE (Root Mean Square Error).

### Шкала баллов

$\text{Points} = 100 \times \max(0, 1 - \text{RMSE}/0.30)$

При  $\text{RMSE} > 0.30$  — 0 баллов. «Точное» решение:  $\text{RMSE} \leq 0.05$ .

## Система оценки

$\text{Points} = 100 \times \max(0, 1 - \text{RMSE}/0.30)$

При  $\text{RMSE} > 0.30$  — 0 баллов. «Точное» решение:  $\text{RMSE} \leq 0.05$ .

## Замечание

### Рекомендации по обработке

- **Предобработка:** Примените CLAHE для усиления контраста, `RandomPerspective` и другие аугментации (например, `RandomRotation`, `ColorJitter`).
- **Базовый baseline:** Fine-tune `EfficientNet-B0` (pretrained на ImageNet21k) с MSE-loss (PyTorch).
- **Постобработка:** Используйте TTA (Test-Time Augmentation, 8 прогонов) и калибровку предсказаний изотонической регрессией.