Контроль качества: автоматическое выявление поддельных товаров

1 сентября 2025 г.

Проблема и бизнес-ценность

- Контрафакт $\rightarrow$ финансовые потери, репутационные риски, санкции.
- Где используем модель: модерация листингов, пост-проверки, выборочный аудит.
- Цели внедрения:
  - **F1 на проде**: *целевое значение*;
  - **Снижение ручной модерации**: *целевой %*;
  - **Скорость проверки / листинг**: *целевое время*.
- Метрики успеха и точки интеграции в процесс.

Визуал: мини-схема бизнес-процесса «до/после», 2–3 KPI.

Постановка задачи

- Бинарная классификация: **original (0)** / **counterfeit (1)**.
- Объект предсказания: товар (`id`).
- Выход: `submission.csv` формата `id,prediction` (0/1).
- Оценка: **F1-score** на скрытом тесте.
- Ограничения: время инференса, размер модели, интерпретируемость.

Данные и модальности

- Табличные признаки: категории, бренд, цена, продавец, метрики листинга.
- Текст: заголовок, описание, атрибуты (NLP).
- Изображения: 1–N фото/листинг (CV).
- Разметка: целевая метка `resolution` только в train.
- Сложности: дисбаланс классов, дубликаты, шумная разметка.
- Политика сплита: без утечек (группы по seller/brand/item).

Визуал: три иконки модальностей с краткими примерами.

Архитектура решения (обзор)

- Препроцессинг по модальностям $\rightarrow$ энкодеры (CV/NLP/Tabular).
- Калибровка вероятностей + подбор порога по F1.
- Кэширование эмбеддингов, логирование, мониторинг.

Визуал: блок-схема пайплайна с потоками данных.

Подготовка данных

Табличные (Tabular)

- Очистка, нормализация.

Изображения (CV)

- Resize, Crop, Flip, Color Jitter; нормализация.
- Контроль качества, near-dup детекция.

Текст (NLP)

- Нормализация, токенизация, max_len/усечение.
- Мультиязычность, спецсимволы.

Дисбаланс

- Class weights / oversampling / focal loss.

Визуал: примеры аугментаций; таблица «до/после».

Базовые модели по модальностям

- Tabular: HistGradient Boost + OOF
- Text:
- Vision: CLIP
- Выход веток: эмбеддинг фиксированного размера + логит.
- Критерии выбора: баланс качества/скорости/ресурсов.

Визуал: три карточки моделей с ключевыми спеками.

Фьюжн / объединение представлений

- Подходы: **late** (взвешенные логиты), **mid** (конкат эмбеддингов → MLP), **stacking** (OOF → мета-модель).
- Выбор: **mid-level фьюжн** — $[e_{cv} \| e_{txt} \| e_{tab}] \rightarrow$ MLP/Attention.
- Маскирование при отсутствии модальности; нормализация.
- Регуляризация
- Абляции: вклад модальностей, чувствительность к пропускам.

Визуал: слой фьюжна с размерами векторов.

Валидация и калибровка

- CV: Stratified K-Fold + GroupKFold (seller/brand) для защиты от утечек.
- Метрики: F1, precision/recall, PR-AUC по фолдам.
- Порог: подбираем по максимальному F1 на OOF; калибровка (Isotonic/Platt/Temp).
- Стабильность: разброс по фолдам, bootstrap доверительные интервалы.

Визуал: PR-кривая, график F1 vs threshold.

Результаты (оффлайн)

- Сводка по фолдам: mean$\pm$std F1; лучшая конфигурация.
- Матрица ошибок; вклад модальностей (CV/TXT/TAB/фьюжн).
- Производительность: tps/batch, VRAM/память, время на 1000 товаров.

Визуал: бар-чарт вкладов, confusion matrix, таблица фолдов.

Анализ ошибок

- Типичные FP: «дешёвый оригинал», редкие бренды, агрессивные аугментации.
- Типичные FN: высококачественные подделки, вводящий в заблуждение текст.
- Кейсы (1–2 примера): картинка + краткий разбор факторов.
- Гипотезы улучшений: OCR, доп. признаки (гео/возраст аккаунта), hard-negative mining.

Визуал: галерея 2×2 (FP/FN) с подписями.

Инференс и выпуск submission

- Единый запуск:

CLI

```
python run.py -mode infer -test_dir data/test \
  -out submission.csv -cfg config.yaml
```

- Входы: пути к данным 3 модальностей; batched инференс; чекпоинты.
- Выход: детерминированный `submission.csv` + логи/метаданные.
- Обработка пропусков модальностей; graceful degradation; мониторинг.

Визуал: фрагмент `submission.csv` (3–5 строк).

Инфраструктура и воспроизводимость

- Docker (CUDA/CPU), фиксированные seed, lock-файлы зависимостей.
- Конфиги: YAML (пути, гиперпараметры, режим фьюжна), единый entrypoint.
- Трекинг: MLflow/W&B; артефакты: модели/эмбеддинги/логи.
- DVC/Git LFS для данных; CI: линтеры, unit/integration, smoke-инференс.
- Модель-реестр, версия API, rollback-стратегия.

Визуал: диаграмма Dev $\rightarrow$ CI $\rightarrow$ Registry $\rightarrow$ Prod.

Риски, ограничения, комплаенс

- Сдвиг данных (сезонность, новые бренды); генерализация на редкие классы.
- Качество снимков/текста; атаки (adversarial, spoofing).
- Этичность и bias; интерпретируемость (SHAP, примеры).
- Легальные аспекты: хранение изображений, PII, retention, аудит.
- Контроль: переобучаемость по расписанию, алерты по PR-AUC/F1.

Дорожная карта и next steps

- Улучшения: HPO (Bayes/ASHA), псевдолейблинг, self-training, CLIP/OCR.
- Инженерия: онлайновая калибровка, микросервис/gRPC, авто-скейлинг.
- Оптимизация: дистилляция, quantization/ONNX/TensorRT.
- Валидация в проде: A/B, holdout-маркировка, human-in-the-loop.

Визуал: дорожная карта (полосы-этапы).