Синаков Арсений Александрович

ML-инженер / Data Scientist

Москва | +7 (991) 314-76-61 | studyfinancee@gmail.com | Telegram: @imballet
github.com/financier-py | financier.su

ОБРАЗОВАНИЕ

НИУ «Высшая школа экономики» *2024 - 2029*
Бакалавриат, совместная программа ФКН и ФЭН «Экономика и анализ данных»

ОПЫТ РАБОТЫ

Сбер - Data Scientist *июнь 2026 - наст. время*
Москва | sberbank.ru

- Спроектировал модель дожития портфеля вкладов с параметрической экстраполяцией кривых удержания на многолетний горизонт. Реализовал батчевую обработку данных на PySpark и Polars, снизив пиковое потребление RAM с 1 ТБ до 128 ГБ. Обучил модель на py_boost с кастомной функцией потерь и маскированием цензурированных исходов.
- Разработал uplift-модели (Meta-Learners, Doubly Robust) для непрерывных таргетов. Написал внутренний модуль валидации поверх EconML: стратифицированная кросс-валидация с контролем баланса тритмента, расчёт AUUC, доверительные интервалы через bootstrap и оценка значимости через permutation-тесты.
- Построил прогноз временных рядов на TFT, N-BEATS и PatchTST с квантильными лоссами. Решение превзошло продакшн-бейзлайн по WAPE и было принято как базовая архитектура.

НИУ ВШЭ - учебный ассистент *сентябрь 2025 - наст. время*
Москва | hse.ru

- Веду несколько групп по курсам «Дискретная математика», «Программирование на Python», «Программирование на C++»: консультации и проверка домашних работ.
- Принимаю устные коллоквиумы, составляю задачи для контестов и экзаменов.

ХАКАТОНЫ И СОРЕВНОВАНИЯ

Ozon E-CUP 2026 - ценность пользователей поиска Polars | LightGBM | CatBoost | Optuna
Прогноз GMV 250 тыс. пользователей маркетплейса на 30 дней вперёд, метрика RMSLE.

- Сгенерировал 300+ признаков на Polars: агрегаты по временным окнам, переходы из поиска в заказ, out-of-fold target encoding.
- Настроил time-split валидацию с проверкой ковариатного сдвига через adversarial validation, признаки отбирал по permutation importance.
- Оптимизировал кастомный RMSLE-лосс в Optuna и собрал двухуровневый стек (LightGBM, CatBoost, XGBoost) на OOF-предсказаниях, снизивший RMSLE на 18% относительно бейзлайна организаторов.

Tender Hack 2025 - RAG-сервис для портала госзакупок
Классификация пользовательских интентов, семантический поиск по векторному индексу нормативной базы и генерация ответов локальной LLM.

ИЗБРАННЫЕ ПРОЕКТЫ

Чат-модератор для русского языка PyTorch | Char-CNN | multi-label [github]
Посимвольная свёрточная сеть для многоклассовой модерации текста без словарной токенизации, устойчивая к обфускации и опечаткам. Метрики: F1 = 0.96 (спам), 0.89 (токсичность), 0.74 (ненормативная лексика).

Деградация глубоких сетей PyTorch | ResNet | BatchNorm [github]
Исследовал затухание градиентов в сетях глубиной от 2 до 40 слоёв: сравнил MLP и ResNet без нормализации, с pre-norm и post-norm BatchNorm и показал, как residual-связи и место нормализации убирают деградацию.

Посимвольное исправление опечаток PyTorch | CNN + Highway + BiLSTM [github]
Модель исправления ошибок на уровне символов, обученная на Википедии и Либрусеке с генерацией опечаток на лету: соседние клавиши, пропуски, перестановки, фонетические замены.

Transformer с нуля и BPE-токенизатор PyTorch | from scratch [github]
Реализовал Transformer (encoder-decoder) с нуля на PyTorch и написал для него свой BPE-токенизатор.

ПУБЛИКАЦИИ

Случайные совпадения теория вероятностей [статья]
Почему совпадения случаются чаще, чем кажется.

ТЕХНИЧЕСКИЙ СТЕК

Языки программирования	Python, C++, SQL
ML / DL	PyTorch, Hugging Face, scikit-learn, CatBoost, LightGBM, XGBoost, py_boost
Модели и методы	Transformer, TFT, N-BEATS, PatchTST, Char-CNN, LSTM, Survival Analysis, Uplift / Causal ML (EconML, CausalML), A/B-тесты, CUPED
Данные	PySpark, Polars, pandas, Hadoop, Hive
MLOps и инструменты	Docker, Git, MLflow, Optuna, Weights & Biases
Иностранный язык	английский, уровень B2