

3+ Years of Experience

Lead Data Scientist

Scoring

Обязанности

- **Лидирую 4 аналитиков:** менторство, обучение, PR. Организация Down Stream 70% (поставка) и Up Stream 30% (приоритезация). Работа над улучшением производительности. Выстраивание коммуникации команды с заказчиками. Найм. Процессы.
- **Работаю с заказчиками:** узнаю проблемы бизнеса, предлагаю решения, оцениваю сложность работы, презентую результат поставки фич на бой, продумываю метрики.
- **Ищу точки роста ключевых метрик:** нахожу проблемы на production, ищу несоответствие логики между моими сервисами и сервисами других подразделений.
- **Оптимизировал процессы аналитической команды:** ускорил поставку типовых задач, улучшил структуру витрин и упростил работу с кодовой базой, что снизило нагрузку на команду и повысило стабильность релизов.

Data Scientist

Scoring

Обязанности

- **Провожу исследования идей по улучшению продукта:** как упаковать лимит клиента наиболее выгодно? С каким новым источником данных делать интеграцию? Как сделать продукт дороже без потери объемов и качества? В общем смысле, отвечаю на вопросы, какую фичу внедрять, а какую нет.
- **Внедряю логику в Python-сервис подбора продуктов и скоринга юзеров.** В задачи сервиса входит подбор параметров продукта (ставка/грейс-период/сроки т.д.), и оценка риска заемщика с помощью ML-моделей и эвристик (какой лимит выдать заемщику?)
- **Делаю e2e A/B тестирования гипотез:** обсуждаю гипотезы, согласовываю метрики, внедряю (если в моей кодовой базе), контролирую набор семплов, анализирую, презентую результат.
- **Работаю с данными:** готовлю отчеты для бизнеса (AirFlow), пишу признаки для ML-моделей (DVC + GIT), помогаю составить витрины данных (ClickHouse + Kafka/Redis),
- **Ставлю задачи разработчикам:** новые фичи для ML-моделей, доработка витрин, организация Postgres таблиц для нужд аналитиков, обогащая данные клиента офф-лайн.

Достижения

- **Внедрил новый модуль для «слияния» двух разных продуктов на один экран мобильного приложения.** Ценность фичи: юзеру на 1 кнопку меньше до покупки. Основная сложность — синхронизация нескольких продуктовых и технических команд в сжатые сроки. В результате удалось внедрить решение с минимальным временем поставки и высокой надёжностью. Разработал стабильный и высоконагруженный модуль, который успешно используется в продакшене и поддерживается командой без дополнительной нагрузки на разработчиков.
- **Переработал продуктовую стратегию в выдачах с фокусом на улучшение качества трафика и устойчивость дохода.** Результат — рост доходности без потерь ключевых метрик, благодаря изменению лимитных правил и сегментации. Основная сложность: определить качество трафика.
- **Оптимизировал процесс добавления новых настроек продуктов.** Ценность: релизы типовых задач перестали переноситься и стало меньше стресса для аналитиков. Моя роль: написал код, согласовал изменения процессов со смежными командами, презентовал коллегам новый процесс.
- **Добился высокой эффективности команды.** Скорость поставки рядовых продуктовых фич команды от точки принятия обязательств до финальных проверок качества на проде зачастую не превышает 45 дней.

Образование

Магистратура

ITMO University, Saint-Petersburg, Russia

Тема: Big Data and Machine Learning

Диплом: Estimation of the Quantum Dots Size Distribution Function using Machine Learning

Бакалавриат

Southern Federal University, Rostov-on-Don, Russia

Тема: Physics

Диплом: Machine Learning Identification of the architecture of Pt-Cu nanoparticles using pair radial distribution of atoms extracted from EXAFS

Публикации

Science Direct Q1 Journal

The ultimate sensitivity of radial distribution functions to architecture of PtCu bimetallic nanoparticles

Computational Materials Science, Q1, Scopus

https://doi.org/10.1016/j.commatsci.2022.111326

Personal info

Date of Birth

23 March, 2000

Nationality

Russian

Languages

Russian: Native

English: B2

Contacts

Telegram: @vvbark

E-mail: v@vbark.ru

Code Wars: vvbark

Stack

Data

- SQL
- ClickHouse
- PostgreSQL
- Hadoop+PySpark
- DVC
- Kafka / Rebbit

Analytics:

- Python
- Pandas
- Matlotlib
- Scipy + Numpy
- Metabase
- Airflow
- Grafana

ML:

- CatBoost
- MLFlow
- Scikit learn

Management:

- Kanban
- WIP
- Приоритезация
- Декомпозиция
- Ответственность

Code:

- Python
- GIT
- OOP / SOLID
- PyTest
- Code review