

ПАВЕЛ ВАЛОВ

Data Analyst / Data Scientist

Email: valov.pm@gmail.com

Phone: +79111178585 ♦ Telegram: t.me/valovpm

GitHub: bit.ly/3KFv5Ha ♦ Google Scholar: bit.ly/3zp5oDx



ОБО МНЕ

Аналитик с опытом разработки, исследований и аналитики в индустрии и академии. Уверенное владение Python, SQL, анализом данных. Честный, ответственный, целеустремленный, свободное владение английским, отточенное годами преподавания и исследований.

НАВЫКИ И УМЕНИЯ

Языки программирования	Python, SQL, R, C#
Библиотеки и технологии	Python (numpy, pandas, scikit-learn, matplotlib, jupyter, etc.) R (dplyr, ggplot2, kernlab, reshape2, rpart, sqldf, tidyr, etc.) .NET (ADO.NET, Entity Framework, LINQ, WinForms, WCF, etc.)
Анализ данных	Теория вероятностей, Статистика, A/B-тестирование
Базы данных	YT (Yandex Table), ClickHouse, PostgreSQL, MS SQL Server
Business Intelligence	Yandex Nirvana, Yandex DataLens, Apache Airflow, Apache Superset
Операционные системы	Linux (Arch, Manjaro, Ubuntu, Bash, CLI), Windows
Языки	Русский, Английский (C2, свободное владение, академическое письмо)

ОПЫТ

Аналитик <i>Яндекс</i>	Июнь 2023 - Август 2024 <i>Санкт-Петербург, Россия</i>
----------------------------------	---

- Предложил, разработал и внедрил методику расчета CPT Администрирования и CPT Подбора для Центра управления жизненным циклом сотрудников, на основе данных от Группы стратегии и финансов.
- Разработал и внедрил процессы выставления PPU Администрирования на основе данных CPT.
- Дорабатывал процессы Crowd'a, формировал дашборды, используя Nirvana, Groovy, YT, DataLens.

Научный сотрудник <i>University of Waterloo</i>	Сентябрь 2014 - Февраль 2023 <i>Уотерлу, Онтарио, Канада</i>
---	---

- Анализировал данные производительности конфигурируемых программных систем в гетерогенных аппаратных и облачных средах, тренировал модели предсказания на основе данных производительности, используя Python (pandas, scikit-learn, matplotlib), R (tidyr, dplyr, reshape2, ggplot2), Microsoft Azure.
- Проводил лекции, практики, проверял код, разрабатывал скрипты тестирования кода, для курсов: 'Алгоритмы и сложность', 'Введение в Computer Science', 'Объектно-ориентированное программирование', 'Требования к ПО', 'Функциональное программирование', используя: C, Java, Python, Racket (Scheme).
- Разработал и опубликовал 4 метода предсказания производительности конфигурируемых систем, используя модели машинного обучения, представил результаты работы на 3 международных конференциях.
- Провел около 4000 часов образовательной деятельности, включая лекции, практики, семинары.

Научный сотрудник <i>Pratt & Whitney Canada</i>	Сентябрь 2015 - Август 2016 <i>Монреаль, Квебек, Канада</i>
---	--

- Анализировал и модифицировал архитектуру системы, используя шаблоны проектирования, UML, SysML.
- Конвертировал классическую архитектуру программной системы в Software Product Line архитектуру.

Инженер-программист <i>Aller Petfood LLC</i>	Сентябрь 2011 - Август 2014 <i>Санкт-Петербург, Россия</i>
--	---

- Анализировал данные по работе производства, разрабатывал систему производственного документооборота используя: C#, .NET (ADO.NET, WCF, WinForms), Microsoft SQL Server (Core, Reporting Services).
- Спроектировал и внедрил распределенную электронную систему производственного документооборота.
- Автоматизировал сбор данных с производственного оборудования и интегрировал его с системой.

PhD Thesis [bit.ly/3P7ja3L], Master Thesis [bit.ly/3zsyXUG]

University of Waterloo

Сентябрь 2014 - Октябрь 2023

Уотерлу, Онтарио, Канада

Бакалавриат 'Прикладная математика и информатика'

Университет ИТМО

Сентябрь 2010 - Июнь 2014

Санкт-Петербург, Россия

РЕПОЗИТОРИИ, ПУБЛИКАЦИИ, ДОПОЛНИТЕЛЬНЫЕ ССЫЛКИ

АВ-тестирование, повышение квалификации

Karpov.Courses

Сентябрь 2024 - настоящее время

www.karpov.courses

- Анализировал параллельные АВ-тесты, оценивал размер выборки, эмпирические ошибки, доверительные интервалы, проводил коррекцию ошибок методом Холма [bit.ly/45W1LFs, bit.ly/3VvUE1u]
- Анализировал как CUPED уменьшает дисперсию в данных и как влияет на p-value [bit.ly/4fWVCgI]
- Анализировал как удаление разного процента выбросов влияет на мощность [bit.ly/3kTMnGb]
- Анализировал как удаление разного процента выбросов влияет на чувствительность [bit.ly/3HYWJQA]
- Анализировал различные методы расчета доверительных интервалов [bit.ly/4p2wHwn, bit.ly/4mB6PWV]
- Сравнил методы добавления эффекта в синтетических АВ-тестах [bit.ly/4g7PRNC, bit.ly/4g1RLPH]
- Рассчитывал размер экспериментальных групп и MDE для АВ-тестов [bit.ly/47VU17D, bit.ly/4g0hsAd]

Аналитик данных, повышение квалификации

Karpov.Courses

Март 2023 - Май 2023

www.karpov.courses

- Анализ нового алгоритма рекомендаций для ленты новостей на улучшение ключевой метрики (CTR):
- А/В-тестирование для демонстрации ухудшения CTR новым алгоритмом рекомендаций используя: трансформации исходных данных (сглаживание Лапласа, Пуассоновский бутстреп, бакетное преобразование), критерии нормальности (Шапиро-Уилка, Д'Агостино), критерии различия распределений (Т-критерий Стьюдента, U-критерий Манна-Уитни), SQL, ClickHouse, Python (pandas, matplotlib) [bit.ly/3kTMnGb]
- Демонстрация повышения чувствительности ключевой метрики методом линейаризации [bit.ly/3IWRXPQ]
- А/А-тестирование с целью проверки неизменности CTR между группами данных [bit.ly/3L0bmCr]
- ETL-пайплайны для отправки отчетов в ClickHouse и Telegram, используя Apache Airflow, Python, SQL:
- Пайплайн мониторинга и отправки отчета в случае возникновения аномалии в метриках [bit.ly/3ZJUEeD]
- Пайплайн отчета в Telegram о метриках двух продуктов в разных временных срезах [bit.ly/3ZJhd31]
- Пайплайн отчета в Telegram о базовых метриках продукта (DAU, views, likes, CTR) [bit.ly/3mvMCi3]
- Пайплайн отправки отчета в ClickHouse о базовых метриках продукта в разных разрезах [bit.ly/3mrzX8N]
- Дашборды для визуализации и анализа ключевых метрик, используя Apache Superset, ClickHouse, SQL:
- Дашборд анализа причины аномального падения активной аудитории новостной ленты [bit.ly/413UzTW]
- Дашборд анализа различий в поведении 'органических' и 'рекламных' пользователей [bit.ly/400X4VR]
- Дашборд анализа базовых продуктовых метрик ленты новостей (likes, view, CTR, etc.) [bit.ly/4138Tfo]
- Дашборд анализа аудиторных метрик нескольких продуктов (DAU, MAU, WAU, etc.) [bit.ly/3obs2NF]

Научный сотрудник

University of Waterloo

Сентябрь 2014 - Февраль 2023

Уотерлу, Онтарио, Канада

- Разработал и опубликовал метод *аппроксимации и трансфера Парето фронтов конфигураций* между гетерогенными облачными средами (Microsoft Azure), используя Python (pandas, scikit-learn, matplotlib). Репозиторий [bit.ly/3nuIs3p] Статья [bit.ly/3oPyUxk] Видео [bit.ly/3bqLP5W] Слайды [bit.ly/3BFIF9f]
- Разработал и опубликовал метод *генерации и трансфера моделей предсказания производительности* конфигурируемых систем между гетерогенными аппаратными средами, используя R (dplyr, reshape2, ggplot2) Репозиторий [bit.ly/3K8o3sA] Статья [bit.ly/3d0FVsF]
- Разработал и опубликовал тематическое исследование по *сравнению различных моделей предсказания производительности* конфигурируемых систем используя экосистему R (tidyr, dplyr, reshape2, ggplot2). Репозиторий [bit.ly/43dJL7x] Статья [bit.ly/3d3T30d]