

# ПАВЕЛ ВАЛОВ

Аналитик данных / Data Analyst

Phone: +79111178585 ♦ Telegram: [t.me/valovpm](https://t.me/valovpm)

LinkedIn: [bit.ly/3ZiILvJ](https://bit.ly/3ZiILvJ) ♦ Email: [valov.pm@gmail.com](mailto:valov.pm@gmail.com)

GitHub: [bit.ly/3KFv5Ha](https://bit.ly/3KFv5Ha) ♦ Location: Санкт-Петербург, Россия



## ОБО МНЕ

Аналитик данных 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 Библиотеки и технологии

Python, SQL, R, C#  
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 Базы данных Business Intelligence (BI) Операционные системы Языки

Теория вероятностей, Статистика, A/B-тестирование  
Microsoft SQL Server (Core, Reporting Services), ClickHouse, PostgreSQL  
Apache Airflow, Apache Superset, Redash  
Linux (Arch, Manjaro, Ubuntu, Bash, CLI), Windows  
Русский, Английский (C2, свободное владение, академическое письмо)

## ОПЫТ

### Аналитик данных, стажер *Karpov.Courses*

Декабрь 2022 - настоящее время

- Анализировал и визуализировал ключевые метрики и события продукта, строил ETL-пайплайны для отправки отчетов в ClickHouse и Telegram, используя: дашборды, A/B тестирование, Apache Airflow, Apache Superset, Python (pandas, matplotlib, jupyter), Redash, SQL (ClickHouse), Telegram API.

### Научный сотрудник *University of Waterloo*

Сентябрь 2012 - Октябрь 2022  
*Уотерлу, Онтарио, Канада*

- 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 *Pratt & Whitney Canada*

Октябрь 2013 - Июль 2014  
*Монреаль, Квебек, Канада*

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

### Инженер-программист *Aller Petfood LLC*

Сентябрь 2008 - Апрель 2012  
*Санкт-Петербург, Россия*

- 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](https://bit.ly/3P7ja3L)], **Master Thesis** [[bit.ly/3zsyXUG](https://bit.ly/3zsyXUG)]*University of Waterloo*

Сентябрь 2012 - Октябрь 2022

*Уотерлу, Онтарио, Канада***Бакалавриат ‘Прикладная математика и информатика’***Университет ИТМО*

Сентябрь 2006 - Июнь 2010

*Санкт-Петербург, Россия***РЕПОЗИТОРИИ, ПУБЛИКАЦИИ, ДОПОЛНИТЕЛЬНЫЕ ССЫЛКИ****Аналитик данных, стажер**

Декабрь 2022 - Февраль 2023

*Karpov.Courses*

- Анализ нового алгоритма рекомендаций для ленты новостей на улучшение ключевой метрики (CTR):
- A/B-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](https://bit.ly/3kTMnGb)]
- Демонстрация повышения чувствительности ключевой метрики методом линеаризации [[bit.ly/3IWRXPQ](https://bit.ly/3IWRXPQ)]
- A/A-тестирование с целью проверки неизменности CTR между группами данных [[bit.ly/3L0bmCr](https://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](https://bit.ly/3ZJUEeD)]
- Пайплайн отчета в Telegram о метриках двух продуктов в разных временных срезах [[bit.ly/3ZJhd31](https://bit.ly/3ZJhd31)]
- Пайплайн отчета в Telegram о базовых метриках продукта (DAU, views, likes, CTR) [[bit.ly/3mvMCI3](https://bit.ly/3mvMCI3)]
- Пайплайн отправки отчета в ClickHouse о базовых метриках продукта в разных разрезах [[bit.ly/3mrzX8N](https://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](https://bit.ly/413UzTW)]
- Дашборд анализа различий в поведении ‘органических’ и ‘рекламных’ пользователей [[bit.ly/400X4VR](https://bit.ly/400X4VR)]
- Дашборд анализа базовых продуктовых метрик ленты новостей (likes, view, CTR, etc.) [[bit.ly/4138Tfo](https://bit.ly/4138Tfo)]
- Дашборд анализа аудиторных метрик нескольких продуктов (DAU, MAU, WAU, etc.) [[bit.ly/3obs2NF](https://bit.ly/3obs2NF)]

**Научный сотрудник**

Сентябрь 2012 - Октябрь 2022

*University of Waterloo**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](https://bit.ly/3nuIs3p)] Статья [[bit.ly/3oPyUxk](https://bit.ly/3oPyUxk)] Видео [[bit.ly/3bqLP5W](https://bit.ly/3bqLP5W)] Слайды [[bit.ly/3BFIF9f](https://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](https://bit.ly/3K8o3sA)] Статья [[bit.ly/3d0FVsF](https://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](https://bit.ly/43dJL7x)] Статья [[bit.ly/3d3T30d](https://bit.ly/3d3T30d)]