

# Операции с моделями машинного обучения (Machine Learning Operations)

Синонимы: MLOps

MLOps (сокращение от Machine Learning Operations) — это фреймворк, объединяющий принципы DevOps (разработка программного обеспечения и ИТ-операции) со специальными потребностями жизненного цикла моделей машинного обучения (ML-моделями). К таким потребностям относятся прежде всего необходимость хранения не только параметров модели (например, весов нейронной сети), но и наборов данных, при обучении на которых эти веса были получены.

Обучающие наборы данных сами по себе могут быть очень велики. Но требуется хранить еще и их версии, полученные на более ранних этапах разработки. В то время как в DevOps достаточно хранить только версии кода. Кроме этого, если DevOps направлен на интеграцию этапов разработки, тестирования и эксплуатации ПО, то MLOps работает не с программами, а с ML-моделями, для которых этапы обучения и оперирования как правило разделены, а развертывание в промышленной среде вообще может производиться вручную.

Все это приводит к увеличению сроков разработки и введения в эксплуатацию ML-моделей, а также повышению рисков возникновения проблем. Именно эти причины и послужили предпосылками для появления технологии MLOps как адаптации DevOps для ML-моделей. MLOps включает в себя практики, инструменты и процессы, направленные на автоматизацию и оптимизацию их развертывания, мониторинга и обслуживания в производственной среде. Он обеспечивает сквозной рабочий процесс, ускоряющий реализацию проектов машинного обучения, и повышающий надежность использующих их систем (ML-систем).

Под ML-системой (или системой, использующей машинное обучение) подразумевается информационная система, в которой присутствует функционал, реализуемый с помощью ML-моделей. MLOps помогает компаниям автоматизировать задачи и быстро развертывать ML-модели, обеспечивая всем участникам (специалистам по анализу данных, инженерам, ИТ-специалистам) возможность бесперебойной работы, мониторинга и улучшения моделей для повышения точности и производительности.

Рабочий процесс MLOps содержит следующие этапы.

**Понимание бизнеса и данных.** Определяется предметная область разработки ML-системы и измеримые критерии эффективности, которым она должна соответствовать, а также оценивается качество данных, имеющихся в рамках задачи. Цель данного этапа — убедиться в осуществимости проекта.

**Подготовка данных.** Решаются задачи сбора данных, их очистки и стандартизации, выделения и отбора признаков.

**Моделирование.** Производится разработка модели машинного обучения. Как правило рассматривается несколько моделей — кандидатов на внедрение в промышленную эксплуатацию, и выбирается одна из них, в наибольшей степени удовлетворяющая критериям эффективности, сформулированным на первом этапе.

**Оценивание модели (офлайн-тестирование).** Проверяется работоспособность обученной модели на экспериментальных наборах данных, в которые часто намеренно вносятся шумы и искажения с целью оценки устойчивости модели к этим явлениям. Если результаты оценки модели показывают ее несоответствие критериям эффективности, то выполняется корректировка модели. Если результаты оценивания признаются успешными, принимается решение о ее внедрении в промышленную эксплуатацию (развертывании).

**Развертывание модели.** Выполняется интеграция модели в существующую информационную систему в производственной среде компании. Оно включает в себя следующие задачи: определение требований к аппаратному обеспечению, оценку работы модели в производственной среде (онлайн-тестирование), в том числе тестирования восприятия пользователем и удобства использования. Кроме этого разрабатывается план действий на случай сбоев модели и настройку стратегии непрерывного развертывания (т.е. автоматизированного внедрения новых версий модели).

**Мониторинг и поддержка.** Организация непрерывного контроля производительности модели и обеспечение ее поддержки в производственной среде. При работе на реальных данных основным риском является деградация модели. Кроме того, на эффективность модели влияют производительность оборудования и существующий программный стек. Поэтому наилучшей практикой для предотвращения падения производительности модели является организация мониторинга, по результатам которого может приниматься решение о необходимости повторного обучения модели для поддержки ее эффективности.

**Экспериментирование с автоматическим протоколированием.** Проведение экспериментов с построенными ML-моделями с автоматической записью результатов. Эксперименты могут проходить по следующему сценарию. Сначала формируются несколько веток результатом каждой из которых является обученная модель. Затем выбирается критерий качества по которому модели сравниваются друг с другом, и выбирается лучшая.

**Версионирование.** Цель управления версиями — обеспечить полноценный доступ к сценариям и моделям машинного обучения, а также обучающим наборам данных, отслеживая их версии. Новые версии ML-моделей могут появляться, например, при

повторном обучении или дообучении модели на новых данных или с использованием другого алгоритма.

**Обеспечение воспроизводимости.** Воспроизводимость в рабочем процессе машинного обучения означает, что каждый этап обработки данных, построения и развертывания ML-моделей должен давать идентичные результаты при использовании одних и тех же входных данных.

**Контейнеризация.** Упаковка ML-моделей в контейнеры, которые могут быть запущены в любых условиях производственной среды. Под контейнером в данном случае понимается автономный исполняемый пакет программного обеспечения, который включает в себя все необходимое для запуска приложения: код, среду выполнения, системные инструменты и библиотеки, настройки. Контейнеры позволяют упаковать приложение и все его компоненты в единый объект, который может быть запущен на любой поддерживающей его системе.

**Оркестрация контейнеров** — это процесс автоматизации развертывания, управления и масштабирования контейнеров, мониторинг их состояния, управление ресурсами и балансировка нагрузки. Кроме этого система оркестрации осуществляет перезапуск контейнеров при внезапном падении из-за внутреннего сбоя.