

Формат экзамена

Проектирование Систем Машинного Обучения

Михаил Хозин



mipt.x5.ru

Цель:

В составе команды 2-4 человека выполнить и задеплоить ML проект по прогнозированию временных рядов

Необходимо:

Выбрать проект, собрать для него данные, обучить модель, разработать API взаимодействия, запустить на сервере и т.п.

Как пройдет экзамен:

- экзамен в форме защиты проекта с обоснованием выбранных вами решений и рассказом о деталях реализации
- обязательно отразить личный вклад каждого участника команды
- в идеале у вас должен быть поднят сервер на котором будет крутиться ваша модель

+ VCS

+ CI/CD

+ ETL

+ Metrics and monitoring

+ Orchestration

+ Docs and API

Качество модели **не оценивается!**

Это единственное обязательное требование:

Заведите для команды удаленный репозиторий.
Вклад в командную работу будет оцениваться в том числе по коммитам.

Если у вас нет коммитов – вы не работали.
Разделите работу в команде так, чтоб коммиты были у всех.

Попробуйте настроить процесс непрерывной интеграции/деплоймента

Не забудьте написать тесты
на все основные этапы пайплайна

Попробуйте разобраться с pre-commit хуками и процессом деплоймента в github и gitlab

Выберите ту систему, которая вам ближе

**Вы вольны выбрать любую задачу.
Но интереснее всего работать в том случае, когда
ваша модель может получать новые данные и
дообучаться.**

Мы предлагаем вам поработать с временными рядами. Так, чтоб данные непрерывно обновлялись:

- курсы валют
- стоимость газа на бирже
- прогноз погоды

Мы не требуем, чтоб модель была супер точной.
Но будет здорово если вы проверите качество
своих предсказаний на реальных данных.

**Вам нужно обосновать выбранные
метрики качества модели.**

Если говорить о прогнозе спроса, то отклонение
вверх и вниз от истинного значения несут разные
риски для бизнеса.

Подумайте что будет в вашем случае.
Возможно, вы решите отслеживать несколько
метрик.

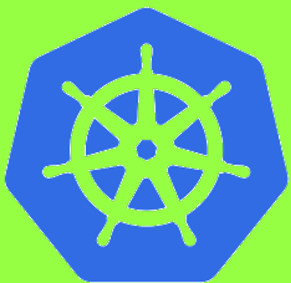
В идеале реализуйте витрину с дашбордами по
вашему проекту.

Разделение вашей задачи на микросервисы

Продумайте как будет организован процесс доставки новых данных в вашу модель

У вас не будет большой нагрузки.
И вам не потребуется поднимать много инстансов.
Но убедитесь, что вы умеете это делать.

Кроме того, убедитесь, что ваш сервис может восстанавливаться после падения.



kubernetes

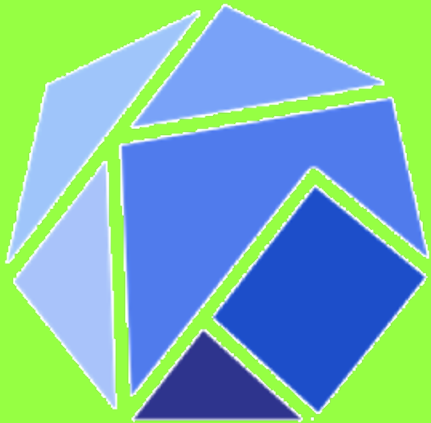
Вам нужно обосновать выбранные метрики качества модели.

Если говорить о прогнозе спроса, то отклонение вверх и вниз от истинного значения несут разные риски для бизнеса.

Подумайте что будет в вашем случае.
Возможно, вы решите отслеживать несколько метрик.

В идеале реализуйте витрину с дашбордами по вашему проекту.

mlflow



Kubeflow

Обсуждать выбранные вами решения НУЖНО!

Вы должны обсуждать решения внутри команды

Но так же вы можете обсуждать выбранный вами путь между командами. Рассказывать о найденных вами решениях, подсказывать друг другу!

Те, кто будет активно помогать сокурсникам и делиться с ними полезными ссылками и подсказками получают плюс в карму и плюс к итоговой оценке.

Не путайте помощь со списыванием!