

Регулярное переобучение

Обучения модели обнаружения мошенничества

Домашнее задание № 5

Цель домашнего задания: отработать шаги, продемонстрированные на вебинаре, включая использование **PySpark** для обучения модели обнаружения мошенничества (*fraud detection*), настройку **MLFlow** с удаленным сервером отслеживания и хранилищем артефактов, сохранение всех артефактов, включая модель, в **MLFlow** с использованием **Object Storage** и настройку переобучения модели по расписанию с помощью **Airflow** на новых данных.

Результаты домашнего задания позволят Вам получить практический опыт в области обнаружения мошенничества с использованием **PySpark** и **MLFlow**. Вы научитесь применять эти инструменты для обучения модели и отслеживания результатов. Кроме того, настройка переобучения модели по расписанию с помощью **Airflow** поможет автоматизировать процесс обновления модели на новых данных, что является важным аспектом в задачах обнаружения мошенничества. Все это позволит развить навыки в области машинного обучения и применения инструментов для разработки и развертывания моделей.

Необходимо на практике отработать шаги, которые были продемонстрированы на вебинаре:

1. Выполнение шагов, продемонстрированных на вебинаре, с использованием **PySpark** для обучения модели обнаружения мошенничества (*fraud detection*);
2. Поднять MLFlow (варианты поднятия на усмотрение студента, рекомендуется использовать сценарий номер 4. Подробнее об этом [тут](#));
3. Сохранить все артефакты с моделью в MLFlow, обязательно наличие Object Storage;
4. Настроить переобучение по расписанию при помощи Airflow на новых данных;
5. Проверить качество и эффективность модели обнаружения мошенничества

Формате сдачи: обновленная ветка main в github репозитории

С вопросами обращайтесь в чат курса.

Общий балл будет рассчитываться как сумма баллов за каждый критерий оценивания.

"Принято" - задание выполнено полностью.

"Возвращено на доработку" - задание не выполнено полностью.

Удачи!