

Методика преподавания машинного обучения с использованием GUI SMILE
Покатов С. П. (ИТМО), Березовская С. В. (ИТМО), Фролова Д. Е. (ИТМО),
Научный руководитель – Доцент Высшей школы цифровой культуры, Романов А.А.
(ИТМО)

Введение. SMILE (Simple Machine Learning Editor) – платформа, позволяющая строить алгоритмы машинного обучения с помощью визуального редактора. Проект SMILE был создан Национальным Центром Когнитивных Разработок Университета ИТМО. Идея платформы была основана на создании высокоуровневого “low-code” инструмента в сфере анализа данных. Сервис позволяет обрабатывать и преобразовывать данные, строить модели машинного обучения, визуализировать данные, оценивать качество построенных моделей, проводить валидацию моделей.

Кроме SMILE существуют также и другие аналогичные ресурсы, позволяющие оптимизировать процессы построения и применения моделей ML за счет простого и понятного графического интерфейса, такие как FEDOT, RapidMiner, KNIME, Yandex DataSphere, Google AutoML

Основная часть. Проект выполняется на платформе SMILE (Simple Machine Learning Editor). Это облачный сервис, который позволяет использовать алгоритмы машинного обучения при помощи визуального редактора. Алгоритмы, лежащие внутри платформы, реализованы на python. Архитектура системы представляет собой ядро, которое может взаимодействовать с различными библиотеками и интегрировать их в одно пространство. Ключевой особенностью является то, что использование платформы не требует навыков программирования. Построение моделей реализуется с помощью визуальных графовых структур, состоящих из узлов и однонаправленных ребер.

Узлы представлены двумя типами – узлы данных и узлы модели. Узлы данных – это сами данные, узлы модели – это различные математические алгоритмы, включающие алгоритм кластеризации, алгоритмы, используемые для задач регрессии и классификации, ансамблевые методы, алгоритмы для работы с временными рядами, предобработка данных, алгоритмы валидации моделей, операции с датафреймами, алгоритмы нейронных сетей, оценка качества моделей. Для каждого алгоритма можно более точно подобрать параметры в настройках узла. Для соединения узлов между собой существуют ребра также двух типов. Обычные ребра и ребра типа apply, необходимые для применения уже обученной модели. Таким образом, можно шаг за шагом, добавляя узлы и соединяя их ребрами, построить пайплайн, - последовательность действий в рамках построения моделей машинного обучения. При необходимости код построенного пайплайна можно выгрузить.

В рамках проекта был построен пайплайн по построению следующих моделей регрессии с предварительной предобработкой данных. Были использованы и опробованы следующие методы: Линейная регрессия, РСА (Метод главных компонент), Логистические регрессия, кластеризация. Реализация не всех методов была корректной. В процессе работы были выявлены недостатки платформы, составлены требования к разработчикам и отправлены на доработку.

Выводы. SMILE является довольно перспективным инструментом. Построение модели с помощью визуального редактора значительно сокращает время работы при условии отсутствия хороших навыков программирования и позволяет решать типовые задачи без использования кода. Такой инструмент может сделать машинное обучение более доступным для специалистов различных областей, чья деятельность не связана напрямую с ML.

Список использованных источников:

1. Acito F. Predictive analytics with KNIME //Analytics for citizen data scientists. Switzerland: Springer. – 2023.
2. Kotu V., Deshpande B. Predictive analytics and data mining: concepts and practice with rapidminer. – Morgan Kaufmann, 2014.