Database-Agnostic Machine Learning

Viction

2024-01-20

Table of contents

Do you have a database? Do you want to make predictions without the hassles of having data scientists? I got you covered!

The Challenge with Relational Data

Most companies store their data in relational databases. These are systems that organize information into tables, like spreadsheets, which can be linked together using keys - like connecting a "Customers" table with an "Orders" table. While this structure is great for storing large amounts of data efficiently, it can be challenging when you want to use machine learning to make predictions or gain insights.

Traditionally, machine learning models expect all the data in a single, flat table. This means analysts have to manually combine tables, calculate summary statistics, and create new "features" from the raw data. This is a process that is very time consuming, error-prone, and specific to each database. If a database changes even slightly, the model may break, and all the work has to be redone - fyi job of the data scientist.

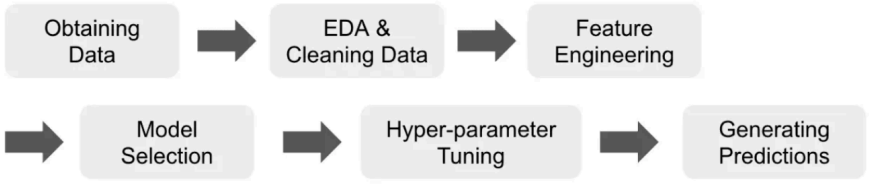

Figure 1: Data Science Pipeline