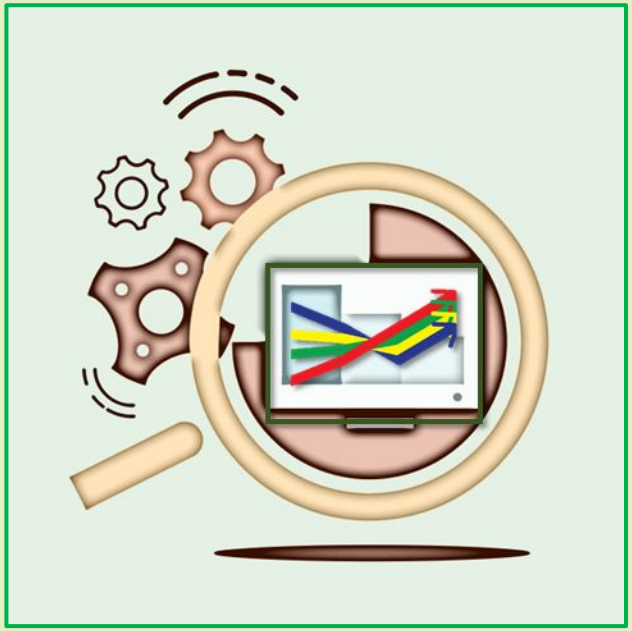
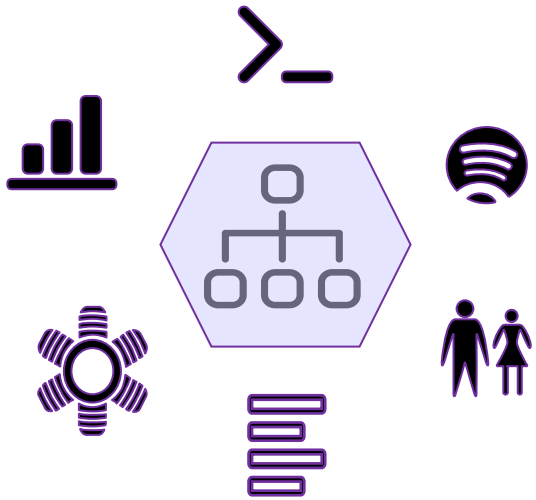
Train In Data



Architecture of our ML API + Implications

ML System Architectures

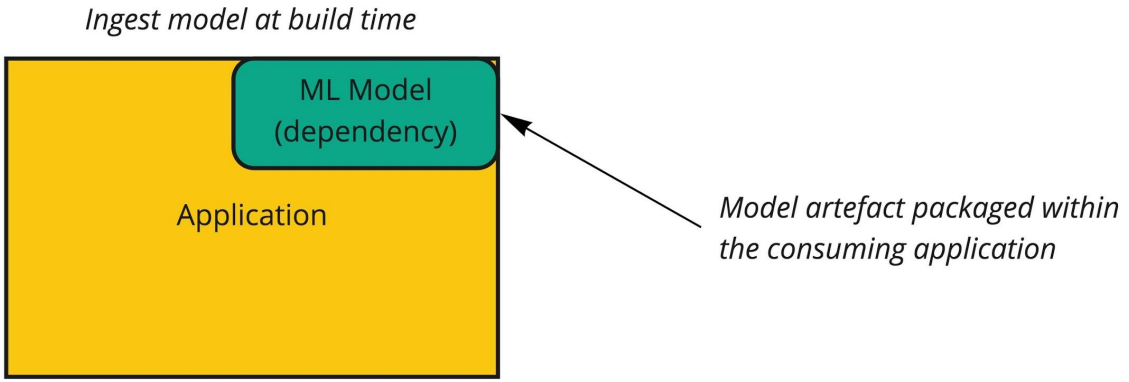


1. Model embedded in application
2. Served via a dedicated service
3. Model published as data (streaming)
4. Batch prediction (offline process)

Architecture 1: Embedded

Pre-Trained: Yes

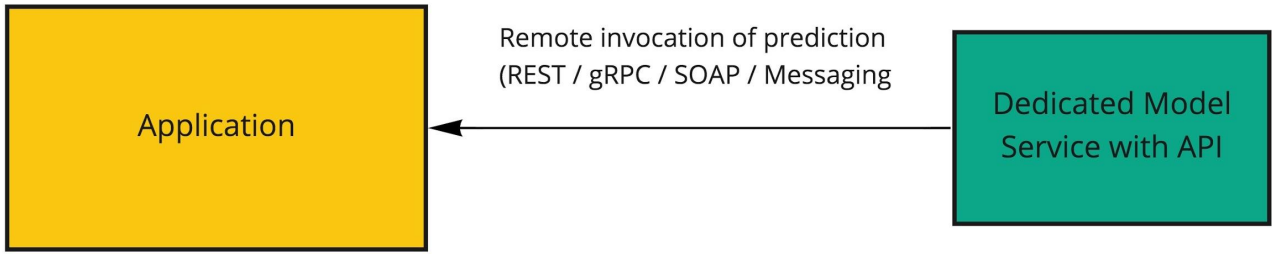
Predict-on-the-fly: Yes



Architecture 2: Dedicated Model API

Pre-Trained: Yes

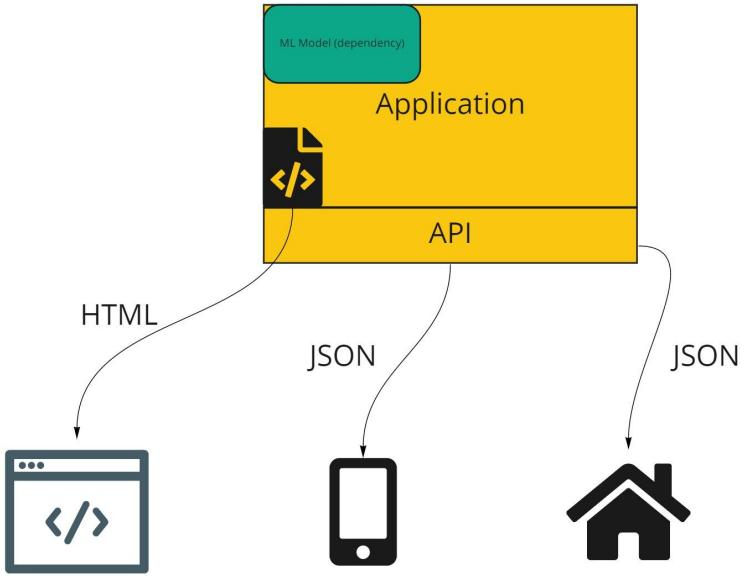
Predict-on-the-fly: Yes



Model is wrapped in a service that can be deployed independently

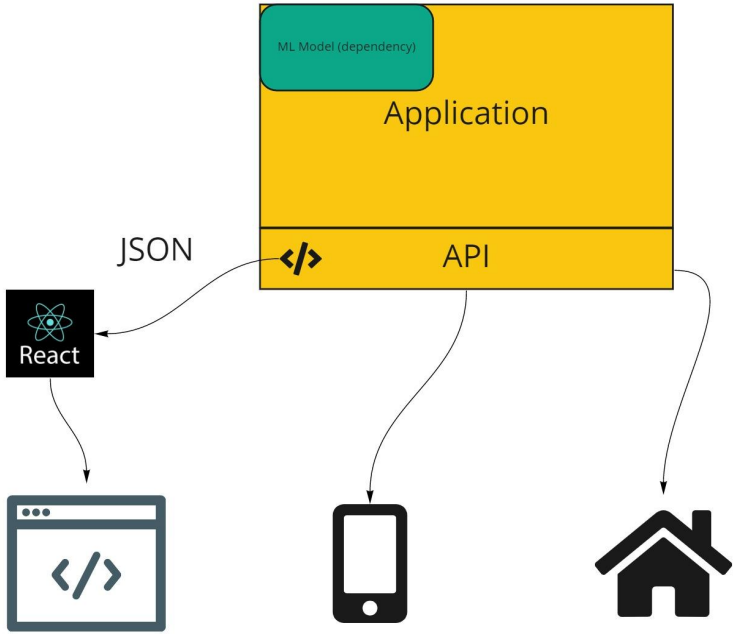
How Can Our API Be Consumed?

- Web browsers (HTML)
- Mobile Devices
- IOT
- Other applications



Modern Frontend Approaches

JS Frameworks like ReactJS, Vue and AngularJS



Dedicated ML API + Microservices

Refer to section 3 for a discussion of the trade-offs

Our example project is not technically "dedicated" - but it's close.