Wie SGKB mit MLOps von omega-ml die KI-Betriebskosten senkt und die Model Governance stärkt

Projekt

Die St. Galler Kantonalbank (SGKB) setzt vielfältig Machine Learning ein, um ihre Kunden kompetent, zur richtigen Zeit am richtigen Ort zu beraten. Mit MLOps von omega-ml senkt SGKB die Zeit für das betriebliche Deployment von Machine Learning-Modellen auf wenige Minuten. Mit dieser Architektur-Standardisierung werden die hohen Anforderungen an Sicherheit und Datenschutz für jeden ML-unterstützten IT Service systematisch eingehalten, die Prognosen werden laufend protokolliert, zentral gespeichert und überwacht.

Herausforderung

Aktuelle und künftige Machine Learning Modelle sollen ohne Zusatzaufwand über den Application-Integration-Layer zur Verfügung stehen, um personalisierte Services in einer Vielzahl von Bank-Anwendungen anzubieten. Bis zur Einführung von omega-ml wurde für jeden Use Case eine eigene technische Schnittstelle entwickelt, was jeweils eine zeitintensive Koordination und aufwändige technische Anpassungen erforderte.

Lösung mit omega-ml

Die omega-ml MLOps Plattform ist in die IT Architektur der SGKB integriert und erfüllt die Anforderungen an eine moderne, skalierbare und sichere Software-Architektur. omega-ml reduziert den Engineering-Aufwand für das Bereitstellen sowie den Betrieb aller ML-Modelle signifikant. Die Nachvollziehbarkeit ist für jedes Modell systematisch gewährleistet.

Standardisierung Das Data Analytics Team deployt ML Modelle und Dashboards innert Minuten als registriertes IT Asset über den Standard-CICD Prozess

Effizienter Betrieb Als standardisierte 24x7 Umgebung für das Modell-Serving und für Analytische Web-Apps stellt omega-ml dedizierte, OpenAPI-kompatible REST APIs und Dashboards im SGKB Kubernetes Cluster bereit.

Monitoring omega-ml erfasst jeden Modell-Aufruf im integrierten Audit-Log, welches für die Drift-Analyse und zur Qualitätssicherung der Modelle genutzt wird.

Modell-Versionierung Neu trainierte Modelle werden von omega-ml versioniert und in JFrog Artifactory abgelegt.

IT-Operations Die Modell-Services sind vollständig in die Operations-Prozesse integriert, über SSO/Keycloak abgesichert und horizontal dynamisch skalierbar.

Nutzen

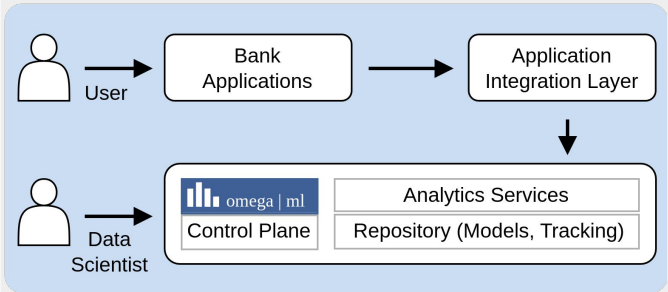
➜ **Nahtlose IT-Integration** Die ML Modelle sind als Microservice über REST API/OpenAPI ansprechbar, gesichert und auditierbar.

➜ **Echtzeit-Anbindung** Die ML-Modelle nutzen die aktuellsten verfügbaren Daten und passen die Prognosen live an.

➜ **Effiziente Deployment-Prozesse** Die CICD Integration ermöglicht Self-Service von Entwicklung bis Produktion und garantiert Nachvollziehbarkeit.

➜ **Höhere Produktivität** Data Scientists entwickeln und optimieren ML Modelle, Analytics-Apps und deren REST APIs ohne Abhängigkeit zu anderen Software Engineers.

Auf einen Blick



Technologie

omega-ml ist als skalierbare MLOps Plattform ausgelegt und mit Plugins erweiterbar. Über die integrierte Control-Plane verwaltet das Analytics Team die Umgebungen *Sandbox* und *Live* für Entwicklung und Betrieb der Modelle.

Python Programmiersprache für ML/AI
XGBoost ML Framework, weitere im Einsatz
RabbitMQ Skalierbare omega-ml Runtime
MongoDB Modelle, Metadaten, Experiment Metrics, Live-Tracking (Drift Analyse, Audit)
SQL Server Data Warehouse, Features-Store
Kubernetes Private-Cloud Deployment
Keycloak IdP-Integration mit JWT