

DATA LAB FRAMEWORK BY INFOMOTION

HÖHERE EFFIZIENZ DURCH REUSEABLE CODE

Das INFOMOTION Data Lab Framework beinhaltet eine Reihe an vorgefertigten Lösungen für Advanced Analytics-Anwendungen. Es bildet alle Schritte einer Data Science Pipeline ab und beinhaltet häufig verwendete Algorithmen und Funktionen in einem Projekt.

HÄUFIGE INEFFIZIENZEN DURCH REDUNDANZ IM CODE

In Advanced Analytics-Projekten treten oftmals wiederkehrende Code-Teile auf, die von Projekt zu Projekt wiederholt übernommen werden. Häufig sind die Pipelineschritte des Data Input, Data Preparation und Feature Engineering betroffen. Diese Schritte bilden u. a. die Anbindung an unterschiedliche Quellsysteme, den Aufbau von unterschiedlichen Aggregationen oder auch die Bereinigung der Daten ab. Häufiges Kopieren von Code sowie redundante, aber gleiche Konfiguration sind fehleranfällig und ineffizient.

Advanced Analytics-Projekte müssen zunächst auf Machbarkeit geprüft werden. Durch statistische Unsicherheiten schrecken Stakeholder häufig von diesem initialen Investment ab. Doch je effizienter die Machbarkeitsstudie, desto kleiner das Investitionsrisiko. Aber auch der Übergang von einem Laborprojekt in die Industrialisierung kann sehr aufwendig sein, da die Pipeline häufig neu entwickelt wird. Denn im Gegensatz zu einem Laborprojekt müssen im produktiven Produkt Standards eingehalten werden und die Software wartbar und stabil sein.

Das Data Lab Framework by INFOMOTION bietet ein standardisiertes Framework an, um durch Abstraktion und Reuseability, also Wiederverwendbarkeit, effiziente Machbarkeitsstudien mit hoher Qualität durchzuführen. Außerdem ermöglicht es durch einen hohen Code-Standard, zentrales Logging, Unit-Tests und weitere vorimplementierte Funktionalitäten eine schnellere Industrialisierung.

DIE KERNSCHRITTE EINER DATA SCIENCE PIPELINE

Das INFOMOTION Data Lab Framework enthält alle wichtigen Pipeline-Schritte für ein Advanced Analytics-Projekt. Typische, wiederkehrende Komponenten sind bereits implementiert. Durch die Ergänzung der projektspezifischen Anforderungen wird aus dem Framework in wenigen Schritten eine vollständige, lauffähige Pipeline.



DATA LAB FRAMEWORK BY INFOMOTION

HÖHERE EFFIZIENZ DURCH REUSEABLE CODE

DATA LAB FRAMEWORK PIPELINE

Das Framework setzt hohen Wert auf Coding und Software-Engineering-Standards (PEP8, Coding Guidelines, etc.). Um die Qualität sicherzustellen, sind für jedes Modul Unit-Tests bereitgestellt. Aber auch die Code-Dokumentation und das Logging sind Kernteile des Frameworks.

1. DATA EXPLORATION

Grundlegende statistische Analysen und Visualisierungen der Daten sind bereits implementiert und können direkt verwendet werden. Eine individuelle Analyse der Unternehmensdaten wird in diesem Schritt projektspezifisch ergänzt.

2. DATA PREPARATION

Die Anbindung von unterschiedlichen Datenquellen wie Datenbanken oder Cloud-Systeme sowie einige grundlegende Datenbereinigungen und Konfigurationen, wie z. B. der Umgang mit fehlenden Werten, sind bereits implementiert. Die gezielte Anpassung der unternehmensspezifischen Daten ist zu ergänzen.

3. FEATURE ENGINEERING

Abgeleitet aus vergangenen Projekten wurden typische Feature-Berechnungen bereits implementiert. Durch die hohe Modularität des Frameworks können die Features je nach Projektanforderung angepasst, erweitert und genutzt werden. Um das Feature Engineering zu vervollständigen, werden auf Basis von kundenspezifischem Fachwissen weitere Einflussfaktoren abgeleitet.

4. MODEL TRAINING & EVALUATION

Das Trennen in Training- und Testdaten, unterschiedliche Trainingsalgorithmen, Hyperparametertuning sowie die Bewertung von Feature-Einflüssen und Modellgüte sind bereits implementiert. Die datenspezifische Modellierung für eine gezielte Auswahl der Algorithmen kann je nach Projektumfang und Anforderung noch erweitert werden. Auch AutoML-Lösungen können integriert werden.

5. POST PROCESSING

In diesem letzten Schritt werden die Ergebnisse nachbereitet und visualisiert. Abschließend werden die Ergebnisse kundenorientiert aufbereitet und in dem gewünschten Format dargestellt. Mit der Anwendung des DLFs wird nicht nur die Entwicklungszeit stark reduziert, sondern die Migration in eine Produktionsumgebung einfacher. Unterschiedliche Schnittstellen und die modulare Entwicklung vereinfachen die Integrierbarkeit in eine bestehende Architektur.

MIT INFOMOTION ZUM ERFOLGREICHEN ADVANCED ANALYTICS-PROJEKT

WIE WEIT SIND SIE?

Mit unserer Lösung wollen wir nicht nur schnelle und effiziente Laborergebnisse erzielen. Wir begleiten unsere Kunden von der Laborumgebung bis hin zur Produktionsumgebung und industrialisieren Ihr Advanced Analytics-Projekt. Haben Sie Interesse an einem Austausch, wie man Ihre Projekte effizienter gestalten kann? Dann kommen Sie gerne für ein unverbindliches Gespräch auf uns zu.

PHILIPP PARAGUYA Manager

INFOMOTION GMBH
Augustinerstraße 10
50667 Köln
T +49 69 56608-3001
phillipp.paraguya@infomotion.de



www.infomotion.de

Die INFOMOTION GmbH ist das führende Beratungsunternehmen für Business Intelligence, Big Data und Digital Solutions im deutschsprachigen Raum. Das Leistungsportfolio reicht von der strategischen Beratung über die Konzeption, Implementierung und den Betrieb nachhaltiger Lösungen bis hin zur Schulung Ihrer Mitarbeiter.

INFOMOTION
we love data