WHITEPAPER

Federated Learning

MIT FEDERATED LEARNING ZU ERFOLGREICHEN KI- UND DATA-SCIENCE-PROJEKTEN

eoda
daten ▸ wissen ▸ nutzen

Hintergründe, Konzept, Use Cases & Realisierung: Die Inhalte im Überblick

1 It‘s (almost) all about the data: Stolpersteine von Datenprojekten

2 Federated Learning: Zusammenführung von Modellen anstatt Daten

3 Federated Learning: Branchenübergreifende Einsatzmöglichkeiten

4 Federated Learning: Die Realisierung

eoda
daten ▸ wissen ▸ nutzen

Federated Learning: Per Paradigmenwechsel zum Datengold

Datenschutz und Datenhoheit als Digitalisierungshürden? Federated Learning kann helfen diese Hürden, vor denen wir alle stehen, erfolgreich zu nehmen.

Federated Learning als Konzept verbindet die Themen Datenschutz und Datenhoheit mit dem Bedarf möglichst umfangreiche Informationen in die Entwicklung und das Training von Analysemodellen einzubeziehen.

Federated Learning kann helfen belastbarere Ergebnisse zu liefern und die Entwicklungszeit von datengestützten Lösungen zu verkürzen.

In vielen Fällen ist Federated Learning gar erst der Schlüssel, um präzise Prognosen überhaupt zu ermöglichen. Mit Federated Learning entstehen Wissensvorsprünge in der Sharing Economy.

Was ist das Konzept hinter Federated Learning? Welche Mehrwerte entstehen für Unternehmen? Was sind erfolgsversprechende Use Cases und welche Schritte umfasst die Realisierung?

Auf diese Fragen geben wir in diesem Whitepaper Antworten. Ich freue mich über Ihr Interesse an diesem zukunftsweisenden Thema, welches die Türen zu KI und Co. für immer mehr Unternehmen öffnen kann.

Viel Spaß beim Lesen.

Herzliche Grüße

Oliver Bracht
Chief Data Scientist bei eoda

1

It‘s (almost) all about the data: Stolpersteine von Datenprojekten

„ Daten sind der Rohstoff des 21. Jahrhunderts.

Der Erfolg von KI- und Data-Science-Projekten ist untrennbar mit der richtigen Datengrundlage verbunden. Genau hier scheitern aber bereits viele Analyseprojekte.

Zentrale Herausforderungen in puncto Datenbasis im Überblick:

1. Datenhoheit

Wem gehören welche Daten? Diese Frage ist in vielen Fällen nicht eindeutig geregelt. Die Verfügungsbefugnis über die Daten ist aber die Grundlage, um sie für Data-Science-Initiativen nutzen zu können. Ein gutes Beispiel für die Komplexität des Themas Datenhoheit ist die Industrie 4.0. Hier ist zu klären, ob die Datenhoheit beim Maschinenhersteller, bei den Herstellern einzelner Komponenten oder dem Betreiber der Anlagen liegt. Ein klarer rechtlicher Rahmen fehlt noch und die Schaffung individueller Regelungen zwischen den Stakeholdern wird notwendig. Unabhängig von den rechtlichen Fragen der tatsächlichen Datenhoheit, ist sie natürlich ein limitierendes Element im Hinblick auf die für die Entwicklung der Algorithmen zur Verfügung stehenden Trainingsdaten.

2. Datenschutz

Der Datenschutz als regulatorische Basis der Datenverarbeitung ist elementar, aber gleichzeitig auch Bremsklotz für das schnelle und umfassende Training von Analysemodellen. Kritisch wird es, wenn es um den Einsatz personenbezogener Daten geht. Neben unmittelbaren personenbezogenen Informationen, wie zum Beispiel Kundendaten, lassen viele weitere

Daten Rückschlüsse auf einzelne Personen und ihre Handlungen zu. Am Beispiel der Industriemaschinen können Maschinenparameter unter Umständen Rückschlüsse auf den bedienenden Mitarbeiter oder die Mitarbeiterin zulassen. Der Datenschutz schränkt dadurch die verfügbare Datenbasis ein und erschwert den Informationsaustausch zwischen Unternehmen.

3. Dezentralität

Viele Unternehmen verfügen nur über begrenzte Datenmengen und selbst diese stehen zum Teil nur einzelnen Abteilungen zur Verfügung. Es entstehen abgeschlossene Datensilos. Datengestützte Mehrwerte entstehen in der Regel aber genau dann, wenn es gelingt unterschiedliche Datenquellen miteinander zu verbinden, um Zusammenhänge aufzudecken.

4. Rare Events

Der vierte Stolperstein basiert auf den drei zuvor genannten und der damit verbundenen oftmals limitierten Datenbasis. Problematisch wird dies insbesondere dann, wenn es darum geht mit Analysen zum Beispiel bestimmte Ereignisse zu prognostizieren. Nehmen wir Maschinenausfälle als Beispiel, dann bilden problematische Verläufe eine starke Minorität. Das Training eines Prognosemodells wird dadurch massiv erschwert und belastbare Ergebnisse bleiben komplett aus oder treten erst mit großem zeitlichen Vorlauf ein.

Datenschutz & KI: Herangehensweisen

Anonyme Daten

- Substantieller Eingriff in Trainingsdaten
- Häufig notwendig, große Datenmengen zu löschen

Synthetische Daten

- Künstlich erzeugte Daten
- Müssen Originaldaten hinreichend gleichen
- Müssen gleichzeitig ausreichend abweichen

Federated Learning

- Daten verlassen nicht die lokale Datenquelle
- Datenzugriff durch Dritte praktisch ausgeschlossen

2

Federated Learning: Zusammenführung von Modellen anstatt Daten

Ein Lösungsansatz für die zuvor geschilderten Herausforderungen: Federated Learning – das föderale Lernen. Federated Learning ist eine spezielle Technik des verteilten maschinellen Lernens, die es ermöglicht, die Modellgüte entscheidend zu verbessern und gleichzeitig die Datenschutzbestimmungen einzuhalten. Wie dies gelingt? Beim Federated Learning wird aus einer Vielzahl einzelner Analysemodelle unterschiedlicher Teilnehmer ein zentrales Modell gebildet. Dieses nimmt wiederum iterativ Einfluss auf die Einzelmodelle.

Das Erfolgsrezept? Die Möglichkeit, eine deutlich umfangreichere Datenbasis für das Training der Modelle einzubeziehen – dezentral und ohne Herausgabe sensibler Informationen.

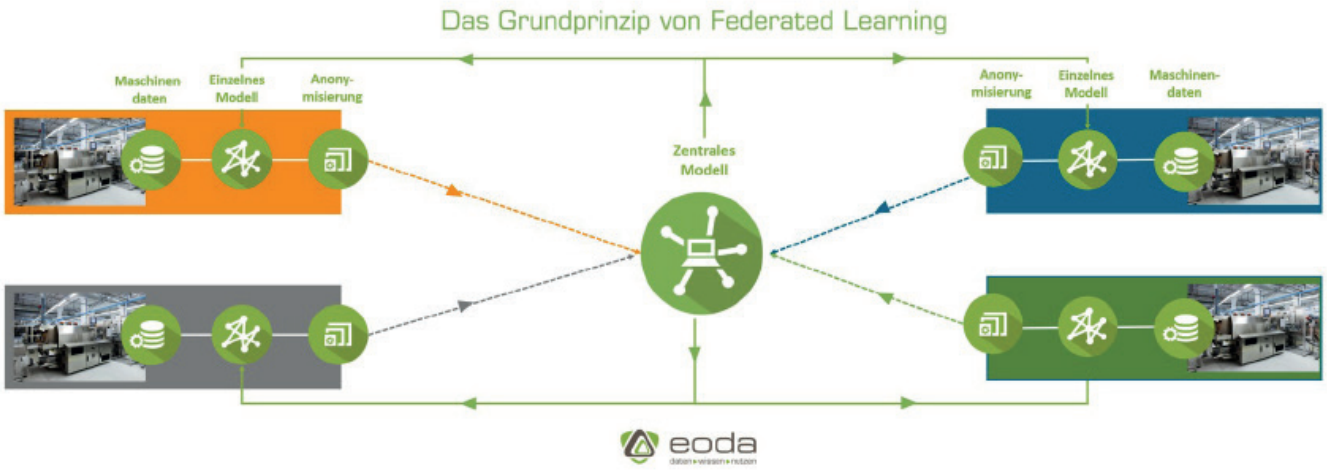
Die Daten verbleiben jederzeit bei den Besitzern. Das zentrale Analysemodell erhält nur die Lernergebnisse, die anonymisierten Parameter, der einzelnen Modelle.

Das Entscheidende: Der entstehende Lerneffekt wird durch die Einbeziehung der Informationen aus dem Training unterschiedlicher Daten massiv verstärkt. Mehrere Modelle werden parallel trainiert und die Genauigkeit der einzelnen Modelle steigt.

Rückschlüsse auf die Ursprungsdaten sind dabei nicht möglich.

Die Vorteile von Federated Learning

- Große Datengrundlage
- Daten verlassen nicht die lokalen Systeme
- Ursprungsdaten bleiben geheim
- Geringere Informationstransfermenge
- Daten können zentral zusammengeführt werden



eoda
daten ▸ wissen ▸ nutzen

3

Federated Learning: Branchenübergreifende Einsatzmöglichkeiten

Industrie & Maschinenbau

Predictive Maintenance ist einer der zentralen Data-Science-Use-Cases in der Industrie. Maschinenausfälle sind dabei häufig sehr seltene Ereignisse mit heterogenem Ursprung. Gerade dieser Kontext ist ein prägnantes Beispiel für die Datenhoheit als Herausforderung. Maschinenbauer oder -betreiber: Wer hat die Hoheit über die Daten und wie können Wege gefunden werden, sodass beide Parteien von den Maschineninformationen profitieren können? Denn: In Ermangelung an einer ausreichenden Menge Trainingsdaten sind präzise Prognosen von Maschinenausfällen in vielen Fällen nicht möglich. Federated Learning kann hier die Lösung sein und das Volumen an Trainingsdaten und damit verbunden auch die Informationen über Maschinenausfälle deutlich erweitern. Anknüpfend an die vorausschauende Instandhaltung kann die Information, wann welches Ersatzteil benötigt wird auch die Lagerhaltung entscheidend verbessern. Predictive Ordering stellt somit die Grundlage für die Vermeidung von Überkapazitäten und Out-of-Stock-Situationen dar.

Federated Learning als Konzept zur Parameteroptimierung kann auch im Bereich der Maschinensteuerung Mehrwerte liefern. Konkret kann sich die optimale Maschinenkonfiguration ermitteln lassen – abhängig von unterschiedlichen Einflussfaktoren, wie dem benötigten Output, den eingesetzten Rohstoffen oder äußeren Einflüssen. Der Algorithmus wird zum Assistenzsystem für die Maschinenführer, die Effizienz der Anlage erhöht und die Bedienbarkeit erleichtert.

So gelingt es den beteiligten Unternehmen Wissensvorsprünge für die proaktive Maschineninstandhaltung zu erzielen, die ihnen als Einzelkämpfer verwehrt geblieben wären.

Gesundheitsweisen

Besonders im Gesundheitswesen sind aussagekräftige und umfangreiche Datensätze aufgrund der Sensibilität der Patienteninformationen schwer zu bekommen. Ein konkretes Beispiel kann hier die Bildverarbeitung von MRT-Aufnahmen zur Tumorerkennung sein. Medizinische Institutionen sind auf ihre eigenen Datenbestände angewiesen, welche aber durch demografische Besonderheiten verzerrt sein können. Federated Learning kann helfen, Erfahrungen aus einer breiten Datenbasis zu sammeln ohne dabei sensible klinische Daten teilen zu müssen. Mittels Differential Privacy kann bei personenbezogenen Daten zusätzlich die Anonymisierung erhöht werden. Hierbei wird ein Rauschen über die Parameterschätzung gelegt um Rückschlüsse auf Einzelpersonen zu verhindern.

Unterhaltungselektronik

Ein typisches Einsatzgebiet für Federated Learning ist die Optimierung von Technologieprodukten im Consumer Bereich. Smartphone-Nutzer haben ein Interesse an der kontinuierlichen Verbesserung ihrer eingesetzten Anwendungen, aber natürlich nicht an der direkten Weitergabe ihrer individuellen Nutzungsdaten. Mit Federated Learning können zum Beispiel die Parameter aus der Analyse der Tastatureingaben von Millionen Nutzern zur Optimierung der Autovervollständigung genutzt werden – ohne, dass persönliche Daten dafür das Smartphone verlassen müssen.

Landwirtschaft

Die Minimierung des Düngemitteleinsatzes, das Monitoring des Pflanzenzustands oder die Steigerung der Prozesseffizienz – Ideen für den Einsatz von Data Science in der Landwirtschaft gibt es viele. Mit Federated Learning können alle landwirtschaftlichen Akteure von datenbasierten Geschäftsmodellen profitieren, ohne dass sie ihre wertvollen Daten dafür preisgeben müssen. Agrarmaschinenhersteller erhalten wertvolle Informationen für die Weiterentwicklung ihrer Maschinen und die Schaffung digitaler Services. Digitalisierte Landwirte profitieren von belastbaren Informationen, die ihnen wirklich helfen können, ihre Arbeit ertragreicher zu gestalten.

Autonomes Fahren

Die Fahrzeugdaten sind die Basis für das Training der KI-Systeme und die Weiterentwicklung des autonomen Fahrens. Diese Daten lassen aber viele Rückschlüsse auf den Fahrer, wie Standort und Aktivitäten, zu. Ein typisches Einsatzszenario von Federated Learning. Die Daten bleiben im Fahrzeug und das zentrale Modell wird mit den Parametern der individuellen Modelle „gefüttert“. So kann die Sicherheit und Zuverlässigkeit autonomer Fahrzeuge erhöht und gleichzeitig dem Datenschutz Rechnung getragen werden.

4

Die Realisierung von Federated Learning

Die Use-Case-Identifikation

Neben den zuvor vorgestellten Einsatzgebieten empfiehlt sich der Einsatz von Federated Learning in vielen anderen Anwendungsfällen. Zu Beginn sollte deswegen die Frage beantwortet werden, welches analytische Problem es zu lösen gilt und inwieweit Federated Learning dabei helfen kann. Entscheidende Faktoren sind hierbei ein hoher Business Value, die grundsätzliche analytische Realisierbarkeit und begrenzt verfügbare Trainingsdaten. Use Cases mit diesen Rahmenbedingungen eignen sich für Federated Learning und beinhalten genug wirtschaftliches Potenzial, damit sich der entstehende Mehraufwand amortisieren kann.

Die Identifikation von Partnern

Kunden, Partner oder Wettbewerber: Wer in Ihrem Umfeld hat einen vergleichbaren Use Case und damit ebenfalls ein Interesse an der erfolgreichen Realisierung und wäre aufgrund der entstehenden Mehrwerte zu einer Kooperation bereit?

Überzeugung der Partner

Nach der Identifikation gilt es die relevanten Partner zu überzeugen und die entstehende Win-Win-Situation zu vermitteln. Hier braucht es natürlich die grundsätzliche Bereitschaft von Zusammenarbeit. Federated Learning ist also auch eine Frage des Mindsets. Wenn Unternehmen die natürlichen Grenzen des Alleingangs in puncto KI und Co. erkennen, können Federated Learning und die Kollaboration sehr attraktiv werden.

Der Abschluss von Verträgen

Ähnlich wie beim Thema Datenhoheit gilt es hier Klarheit zu schaffen und die Rahmebedingungen zum Austausch der Modellparameter zu regeln.

Die Durchführung eines PoCs

Der Use Case und der Einfluss von Federated Learning wird erlebbar. In der Proof-of-Concept-Phase wird das Analysemodell entwickelt und der Anwendungsfall erprobt.

Der Aufbau einer Infrastruktur

Der PoC war erfolgreich und das Analysemodell soll für den Produktiveinsatz ausgerollt werden. Dafür braucht es die richtigen technischen Rahmenbedingungen, um das Analysemodell nahtlos und performant in die Geschäftsprozesse einzubinden.

Der produktive Einsatz

Das Analysemodell ist im Produktiveinsatz. Datengetrieben mit Federated Learning ist aus der Use-Case-Idee ein Datenprodukt geworden, welches eine hohe Relevanz für den Ablauf der angeschlossenen Geschäftsprozesse hat.

Weiterer Roll-out

Das Konzept Federated Learning kann danach auf weitere Use Cases übertragen werden. Auch können weitere Partner miteingebunden werden.

YUNA elements – ein erster Schritt

Ein fiktives Beispiel:

Als Maschinenhersteller bietet man seinen Kunden mit kontinuierlichen Verbesserungen einen Mehrwert. In diesem Beispiel handelt es sich um ein optimiertes Modell zur Vorhersage von Maschinenausfällen sowie zur Planung von Serviceeinsätzen.

Es existieren vier Satelliten-Installationen von YUNA elements in den lokalen Netzen unterschiedlicher Maschinenbetreiber, die sich physisch an unterschiedlichen Orten befinden. Diese Maschinen produzieren Sensordaten, die zur Vorhersage von Maschinenausfällen oder zur Optimierung der Betriebsparameter verwendet werden können.

Die Satelliten-Installationen können nicht miteinander, wohl aber über eine dedizierte Verbindung mit einer Zentralinstanz von YUNA elements des Maschinenherstellers kommunizieren.

1. Satelliten im lokalen Kundennetzwerk

Beim Maschinenbetreiber vor Ort wird eine vom Hersteller vorkonfigurierte und von den Betreibern abgenommene YUNA elements-Installation in die bestehende Infrastruktur installiert. An diesem Punkt besitzt YUNA elements vorgegebene Skripte zum Training eines maschinellen Lernverfahrens, welches für die Prognosen eingesetzt werden soll. Dieses kann mit historischen Daten direkt vor Ort bei den Betreibern trainiert werden.

Das erzeugte Modell kann einerseits sofort vor Ort vom Kunden eingesetzt werden, andererseits dient es der Algorithmenoptimierung durch den Hersteller, indem die Modellparameter an dessen zentrale YUNA elements Instanz übertragen werden. Der Hersteller kann anschließend durch Aggregation der Modelle ein global optimiertes Modell zur Prognose der Maschinenausfälle erzeugen und wieder an seine Kunden verteilen, sodass diese das optimierte Modell produktiv einsetzen können. Dadurch fließen die Erfahrungen aus sehr unterschiedlichen Einsatzbereichen bzw. Betriebssituationen in ein gemeinsames Modell ein.

Die Sensordaten können beispielsweise direkt aus den Maschinen oder aus einer lokalen Datenbank ausgelesen werden. In der lokalen YUNA elements Instanz werden dann die vom Hersteller bereitgestellten Skripte zur Aufbereitung der Sensordaten, der Feature-Extraktion und der anschließenden Modellerzeugung verwendet. Sensible Daten verbleiben dabei im lokalen Netz. Im Anschluss an die Modellerzeugung werden nur die Parameter des trainierten Lernverfahrens an den Hersteller übertragen. Die lokale YUNA elements Installation kann dabei flexibel an die jeweiligen Kundenbedürfnisse angepasst werden. Sensible Daten werden in diesem Modell nur lokal von YUNA elements geladen und bearbeitet und sind von außen nicht sichtbar.

2. Zusammenführen der Modelle in der Zentralinstanz beim Hersteller

In der zentralen YUNA elements Instanz beim Maschinenhersteller werden die übertragenen Kundenmodelle gespeichert und zu einem zentralen Modell zusammengeführt. Jedes Mal, wenn ein Kunde ein aktualisiertes Modell überträgt, kann hier eine Aktualisierung des Zentralmodells angestoßen werden. Im Anschluss wird dieses für die Anwendung an alle Kunden zurück übertragen und kommt dort zum produktiven Einsatz.

Dies hat einen entscheidenden Vorteil: Maschinen bei unterschiedlichen Betreibern besitzen selten dieselben Aufgaben und werden noch seltener mit den selben Parametern betrieben. Durch den Upload eines Modells einer einzelnen Maschine sowie das Einfließen dieser Parameter in das zentrale Modell, kann das Modell stets ein breites Spektrum an Vorkommnissen abbilden. Schließlich wirkt sich dies positiv auf die Wirtschaftlichkeit der einzelnen Maschinen aus.

Beratung, Tools & Umsetzung:

Wir sind Ihr Ansprechpartner

zum Thema Federated Learning.

Ihr Ansprechpartner:

Oliver Bracht

E-Mail: sales@eoda.de

Tel.: +49 561 87948-370

Wir sind Ihr Partner im Umfeld von Big Data, Machine Learning und Künstlicher Intelligenz. Wir unterstützen Sie ganzheitlich – von der Identifikation des richtigen Anwendungsfalls über die Datenanalyse und Interpretation der Ergebnisse bis hin zur Implementierung der entwickelten Lösung in Ihr Produktivsystem.

Effektivere Vertriebskampagnen, zuverlässigere Industrieanlagen oder optimierte Lagerbestände: Der Schlüssel zur Erreichung Ihrer Ziele liegt in den Daten. Wir helfen Ihnen, sich Daten zunutze zu machen und sich im Zeitalter der digitalen Transformation bestmöglich aufzustellen.

Schreiben Sie mit uns Ihre digitale Erfolgsgeschichte.

Mehr Informationen auf www.eoda.de.

eoda
daten ▸ wissen ▸ nutzen