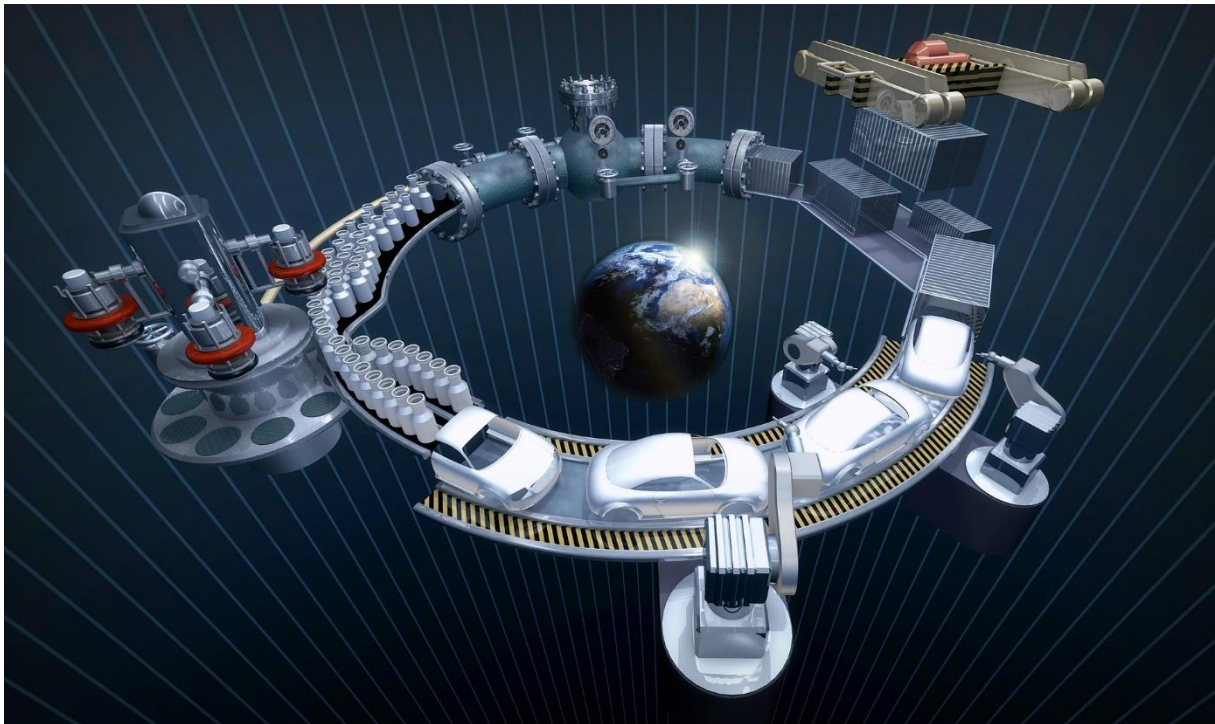


Forschungsprojektantrag für den Forschungsmaster Data Science

Projekttitle: Data Science Plattform für Produktionssysteme



Projektübersicht

Startsemester (von-bis)	WiSe2019/20-SoSe2021
Anzahl Studierende	1
Art (gefördertes Projekt, Projekt mit externen Partnern, Studienprojekt)	Studienprojekt
Projektverantwortung	Prof. Dr. Stefan Berlik
Projektkontext	Projekt in Zusammenarbeit mit dem Center for Applied Data Science (CfADS) in Bielefeld / Gütersloh und der IoT-Factory

Kurzbeschreibung

Die Digitalisierung der Produktionsmittel wie auch der Produkte schreitet immer weiter voran. Sie liefert einen immensen Datenstrom, der als Basis verschiedenster Data Science Anwendungen neue Einsichten und Erkenntnisse ermöglicht.

Im Kernbereich in der Produktion kommen KI basierte Anwendungen beispielsweise bei der intelligenten Instandhaltung zum Einsatz (Predictive- und Prescriptive-Maintenance), um Defekte zu detektieren und vorherzusagen bevor sie Probleme verursachen. Durch das Verhindern ungeplanter Stillstände helfen sie hohe Kosten zu vermeiden. Im Rahmen der Qualitätssicherung können z.B. Dank Bildverarbeitung kleinste Details und Defekte weitaus zuverlässiger erkannt werden als durch das menschliche Auge. Sie steigern die Effizienz und Effektivität des Produktionsprozesses, indem optimale Betriebspunkte angesteuert werden, die den Ressourcenbedarf (Roh-, Hilfs- und Betriebsstoffe) sowie die Durchlaufzeiten senken und so zu geringeren Umweltbelastungen bei höheren Erträgen führen. Bei der Produktionsplanung und -steuerung (PPS) helfen sie, optimale Arbeits- und Ablaufpläne zu finden.

Im weiteren Kontext des Referenzarchitekturmodells Industrie 4.0 (RAMI 4.0) kommen KI basierte Verfahren beim Produktdesign zum Einsatz, wo sie helfen, schnellere und bessere Entwürfe zu generieren. Sie unterstützen beim Post-Production Support, indem sie überwachen, wie Produkte genutzt werden und Klienten diese Informationen zur Verfügung stellen. Beim Kundenservice ermöglichen sie, eine ganzheitliche Ansicht des Kundenerlebnisses in Echtzeit zu generieren, anhand derer Probleme priorisiert und wichtige Kunden und Schwachstellen identifiziert werden können. Ganz allgemein können KI Methoden auch einfach dabei helfen, nützliche, bisher aber ungenutzte Daten zu identifizieren und zu erschließen.

Alle genannten Anwendungsbeispiele erfordern Daten, umfangreiche Daten. Diese stehen in konventionellen Manufacturing Execution Systemen (MES) bzw. Betriebsdatenerfassung (BDE) Systemen aber nicht hinreichend zur Verfügung. Üblicherweise sind hier noch klassische relationale Datenbanken im Einsatz und Data Science (Analytics) Anwendungen werden – wenn überhaupt – über diverse proprietäre Parallelsysteme wie QlikView, Power BI, spezielle Systeme eines Maschinenlieferanten o.ä. abgewickelt. Kennzeichnend für dieses Vorgehen ist das Vorhandensein mehrerer verteilter Datensilos.

Aufgabenbeschreibung des Studierenden

Im Gegensatz dazu soll im Rahmen dieses Projekts eine universale, integrative, offene Data Science Plattform für Produktionssysteme erstellt werden, auf deren Basis als Proof-of-Concept eine KI basierte Applikation aus dem Produktionsumfeld realisiert wird.

Problemstellung und Ziele

Es soll eine Data Science Plattform auf Basis des Data Lake Konzepts erstellt werden, die die Kernfunktionalitäten *data ingestion* (z.B. data streams per MQTT oder OPC-UA) und *data storage* (NoSQL) bereitstellt und möglichst alle Anforderungen erfüllt, die die Anwendung der beschriebenen KI Techniken stellen. Hierzu sind entsprechende Anforderungsanalysen und Use-Cases zu erstellen. Das Systemdesign soll dem Virtualisierungskonzept folgen und entsprechend Techniken wie virtuelle Maschinen, Container und Kubernetes nutzen; auf freier Software aufbauen, auf offene Standards setzen und frei zur Verfügung stehen. Gleichzeitig sollen aber auch Schnittstellen zur Integration in Bestandssysteme geschaffen werden.

Aufbauend auf der beschriebenen Data Science Plattform soll im Rahmen dieses Projekts prototypisch eine Data Science Applikation realisiert werden, vorzugsweise aus dem Bereich Produktionsplanung und -steuerung, Qualitätssicherung oder Produktionsprozessoptimierung. In Absprache mit der / dem Studierenden können allerdings auch alternative KI-Konzepte industrieller Anwendungen realisiert werden.

Bezug zum Thema Data Science

Die erste Projektphase ist schwerpunktmäßig dem Bereich Data-Engineering zuzuordnen, die zweite Phase, der Proof-of-Concept einer KI basierten Applikation aus dem Produktionsumfeld, dem Bereich Data Science.

Benötigte Ressourcen und Sicherstellung der Verfügbarkeit (Daten, Projektpartner, Hard-/ Software)

Die Infrastruktur besteht aus:

- Data-Analytics-Cluster: Mehrere Hochleistungsrechner mit GPU Unterstützung für Deep-Learning-Anwendungen und installiertem Hadoop Big-Data-Framework

- Industrial-IoT-Factory: Modellfabrik als Nachbildung einer realen Produktionslinie vom Rohstoff bis zur Lagerung des fertigen Produkts

Die Datengrundlage ist durch die IoT-Factory gegeben. Bei Bedarf könnten zusätzlich Real-World-Daten aus Produktivsystemen von Industriepartnern genutzt werden.

Grober Projektplan über die 4 Semester

Erstes Semester: Das Ziel des ersten Semesters ist es, sich mit den benötigten Technologien vertraut zu machen. Dazu erfolgt die Einarbeitung in die Themengebiete Big Data, maschinelles Lernen, M2M-Kommunikation, MES, BDE, Produktions- und Logistiksysteme. Die Prüfungsleistung ist das Erstellen eines Forschungsexposés auf Englisch und ein dazugehöriges Kolloquium.

Zweites Semester: Im zweiten Semester soll durch eine Literaturrecherche der aktuelle Stand zu relevanten Arbeiten im Themenfeld ermittelt werden. Im Rahmen einer Anforderungsanalyse sind die durch die Data Science Plattform zu erfüllenden Eigenschaften in Form von Use-Cases zu erfassen. Ziel ist es, diese mit Hilfe der im ersten Semester erworbenen Fähigkeiten und der Erkenntnisse aus der Literaturrecherche in einen Systementwurf der Data Science Plattform münden zu lassen. Dieser soll in Form eines Papers als Prüfungsleistung gewertet und idealerweise im Rahmen einer Konferenz präsentiert werden.

Drittes Semester: Im dritten Semester soll der Systementwurf implementiert werden. Seine Eignung wird durch das Implementieren des prototypischen KI Anwendungsfalls evaluiert. Auch hier ist das Veröffentlichen der Ergebnisse als Paper und eine Präsentation im Rahmen einer Konferenz erwünscht.

Viertes Semester: Inhalt der Masterarbeit ist das Erstellen einer funktionsfähigen Data Science Plattform und eines darauf aufsetzenden prototypischen KI Anwendungsfalls. Dabei sollen die Ergebnisse aus dem zweiten und dritten Semester zusammengeführt werden. Die Ergebnisse werden in der Masterarbeit dargestellt. Ergänzend ist das Schreiben eines Journal-Papers möglich. In jedem Fall wird die Masterarbeit mit einem Kolloquium abgeschlossen

Kriterien zur Überprüfung der Eignung der Bewerberin oder des Bewerbers

Zwingend:

- Bachelorabschluss in einer einschlägigen Fachrichtung (Informatik, Elektrotechnik, Mechatronik, angewandte Mathematik, Kognitionswissenschaft o.ä.)
- Fortgeschrittene Programmierkenntnisse in mindestens einer objektorientierten Programmiersprache, möglichst in Python
- Fließendes Englisch in Wort und Schrift

Optional:

- Erfahrungen mit relationalen und nicht-relationalen Datenbanken
- Grundlegende Kenntnisse zur M2M-Kommunikation
- Grundlegende Kenntnisse im maschinellen Lernen
- Erfahrung mit MES, BDE, Produktions- und Logistiksystemen

Erwerbbarer Kompetenzen durch das Projekt

Die / der Studierende wird mit Abschluss des Studiums über fundierte theoretische und praktische Erfahrungen sowohl im Data Engineering als auch im Data Science Bereich verfügen. Der Erwerb hinreichender Kompetenzen aus beiden Disziplinen ist essenziell für das erfolgreiche spätere Bearbeiten eines jeden Data Science Projekts.

Sie / er ist nach Abschluss des Projekts in der Lage:

- integrierte Big Data Systeme zur Erfassung, Speicherung und Verarbeitung von Daten aufzusetzen und zu betreiben,
- neue Datenströme in das System zu integrieren,
- die im industriellen Kontext relevanten ML-Verfahren zielgerichtet anzuwenden, anwendungsbezogen anzupassen und ggf. weiterzuentwickeln,
- eigene Forschungsergebnisse vor einem Fachpublikum zu präsentieren,
- wissenschaftliche Texte zu verfassen.