Profil

Sebastian Nichtern

Lead Cloud & AI Platform Engineer

Übersicht

Sebastian Nichtern ist Experte für **Künstliche Intelligenz** und **Cloud-Automatisierung** mit über sieben Jahren Erfahrung in der Entwicklung und Produktivsetzung skalierbarer KI-Systeme für Großunternehmen.

Er unterstützt Organisationen dabei, KI-Anwendungen effizient in den Betrieb zu bringen, Prozesse zu automatisieren und datengetriebene Produkte auf Cloud-Plattformen (Fokus AWS) umzusetzen.

Sebastian hat unter anderem:

- Bei **Hapag-Lloyd** produktive **NLP-Systeme** umgesetzt, die über 190.000 E-Mails pro Monat automatisiert verarbeiten (WirtschaftsWoche Best of Consulting Technology Award 2020).
- Bei **Mercedes-Benz** den Aufbau eines unternehmensweiten **KI-Ökosystems und Agent-Gateways** technisch geleitet.
- Für **Vodafone** eine **skalierbare KI-Plattform** zur Standardisierung von ML-Deployments konzipiert und aufgebaut.
- Für **Merck** an einer **Softwareanwendung für KI-gestütztes Wirkstoffdesign** gearbeitet.
- Für die **Deutsche Bahn** eine **KI-Lösung zur frühzeitigen Erkennung von Radschäden** entwickelt und in Pilotbetrieb überführt.

Mit Erfahrung aus Unternehmensberatung und als Gründer mehrerer Technologieunternehmen verbindet er tiefes technisches Verständnis mit strategischem Denken und klarer Kommunikation.

Fokus:

- KI-Automatisierung und datengetriebene Prozessoptimierung
- GenAI-Anwendungen und intelligente Agentensysteme
- Design und Entwicklung sicherer, skalierbarer Cloud-Architekturen (AWS)
- MLOps, produktive ML-Systeme und Betriebsmodelle

Projektübersicht

Jahr (Laufzeit)	Kunde	Projekttitel
2026 6 Monate	**dmTECH**	Evaluation und Konzeption eines agentenbasierten Identitäts- und Berechtigungsmanagements (Agentic IAM) inkl. Zielarchitektur, Anforderungskatalog und MVP-Validierung
2025–2026 15 Monate	**Mercedes-Benz Tech Innovation**	Aufbau eines KI-Ökosystems und Agent-Gateways zur Befähigung interner AI-Teams bei der Umsetzung von GenAI- und agentischen Use Cases
2023 3 Monate	**Leap (Startup)**	Design und Implementierung eines skalierbaren, serverlosen Backends für eine GenAI-Mobile-App auf AWS
2022–2024 24 Monate	**q·ops (Startup)**	Architektur und Aufbau einer skalierbaren End-to-End-MLOps-Plattform auf AWS
2022 1 Monat	**Hapag-Lloyd**	Bewertung von Anbietern und Architekturentwurf einer hochskalierbaren IoT-Datenplattform zur Überwachung von über 3 Mio. Containern
2022 3 Monate	**Hapag-Lloyd**	Reifegradbeurteilung und Konzeption einer AWS-basierten KI-Plattform in Zusammenarbeit mit dem KI-Center of Excellence
2021–2022 4 Monate	**Deutsche Bahn**	Predictive-Maintenance-PoC zur Erkennung von Radschäden mittels Gleissensorik und SAP-Schadcodes
2021 4 Monate	**Vodafone**	Planung und Entwicklung einer KI-Plattform mit Model-Serving-Blueprint zur Standardisierung von ML-Deployments
2021 1 Monat	**Vodafone**	Proof of Concept einer Web-App zur Informationsextraktion (NER) aus Dokumenten mit Human-in-the-Loop
2021 2 Monate	**Merck Group**	Auslegung und Planung einer datengetriebenen, KI-gestützten Applikation zur Wirkstoffentwicklung
2020 3 Monate	**Merck Group**	Bewertung des Potenzials zur Rekonstruktion von Trainingsdaten durch Angriffe auf ML-Modelle
2020 3 Monate	**Hapag-Lloyd**	Deployment eines BERT-basierten NLP-Modells zur E-Mail-Klassifikation im „Dispute Capturing“-Prozess
2020 3 Monate	**Hapag-Lloyd**	Deployment eines NLP-Modells zur E-Mail-Klassifikation im Buchungsprozess der Containerschifffahrt
2019 3 Monate	**Hapag-Lloyd**	Rapid Prototyping und Business-Case-Analyse für 6 NLP Use Cases
2019 2 Monate	**Vossloh**	PoC Anomalieerkennung in Sensordaten von Bahninfrastruktur

Qualifikation und Fachwissen

Methodisches & Strategisches Wissen

- Machine-Learning-gestützte Problemlösung, Systemdesign und Implementierung
- Cloud-native Architekturdesign, Entwicklung, Deployment & Betrieb
- Architektur- und Implementierungsentscheidungen unter realen Produktions-Constraints
- End-to-End Entwicklung und Betrieb von ML-Systemen (Lifecycle, Wartung, Monitoring)
- Qualitäts- und Produktionsreife von Softwaresystemen (IaC, CICD, Teststrategien)
- Klare technische Kommunikation komplexer Systeme

Cloud, DevOps & Infrastruktur (5+ Jahre)

- Cloud Engineering & Solution Architecture Design (AWS)
- Networking & Security (VPC, Netzwerkarchitekturen, Zugriffskonzepte)
- Infrastructure as Code (AWS CDK, Terraform)
- Compute & Orchestration (Lambda, ECS, Fargate, AWS Batch)
- Storage & Data Services (S3, DynamoDB, RDS, AWS Timestream, MongoDB, Firestore, R2...)
- Content Delivery & Monitoring (CloudFront, CloudWatch, CloudTrail)
- DevOps- & Deployment-Strategien (CI/CD, Betrieb, Skalierung)

MLOps & ML Engineering (5+ Jahre)

- Design und Betrieb skalierbarer MLOps-Plattformen
- ML Lifecycle Management (Training, Versionierung, Deployment)
- Monitoring von Modellen (Performance, Drift, Stabilität)
- Datensicherheit beim Bereitstellen von ML-Modellen
- Automatisierte ML-Pipelines und reproduzierbare Experimente

Data Science & Machine Learning (3+ Jahre)

- Explorative Datenanalyse & Datenvisualisierung
- Deep Learning (PyTorch, TensorFlow/Keras)
- NLP & Textanalyse
- Zeitreihenanalyse & Vorhersagemodelle
- Klassifikationsmodelle & prädiktive Analytik
- Computer Vision & Bilderkennung

GenAI & LLM-basierte Systeme (1+ Jahre)

- Entwicklung produktiver Generative-AI- und LLM-Anwendungen
- Agenten- und LLM-Systemdesign (Orchestrierung, Zustandsmanagement)
- AI Governance, Agentic IAM und sichere Betriebsmodelle für KI-Agenten
- Retrieval-Augmented Generation (RAG)

Sprachen

- Deutsch – Muttersprache
- Englisch – Verhandlungssicher (C1/C2)

Beruflicher Werdegang

2024 – Heute	Selbstständiger Lead Cloud und AI Platform Engineer, Hamburg (Remote)
2022 – 2024	Gründer & Geschäftsführer, q-ops GmbH sowie Aufbau von Hello-Menu und Leap, Hamburg
2021 – 2022	Senior Data Scientist, Ginkgo Analytics (heute Eraneos), Hamburg
2019 – 2021	Data Scientist, Ginkgo Analytics GmbH (heute Eraneos), Hamburg

Aus- und Weiterbildung

2017 – 2018	**Weiterbildendes Studium der Informatik** *(Fokus: Data Science & Maschinelles Lernen)* Universität Hamburg
2014 – 2017	**Master of Science (M.Sc.) Verfahrenstechnik** Technische Universität Hamburg *(Note 1,8)*
2010 – 2014	**Bachelor of Science (B.Sc.) Verfahrenstechnik** Technische Universität Hamburg

Auszeichnungen & Förderprogramme

2022 – 2023	Förderung der q-ops GmbH durch das Ideation-Programm des AI Startup Hub Hamburg
2020	WirtschaftsWoche Best of Consulting Award – Technologie (mit Ginkgo Analytics, heute Eraneos)

Kundenprojekte und Referenzen

Evaluation und Konzeption eines agentenbasierten Identitäts- und Berechtigungsmanagements (Agentic IAM) inkl. Zielarchitektur und MVP-Validierung

Kunde: dmTECH, Deutschland
Projektrolle: Solution Architect (AI Governance & IAM), stellvertretender Projektverantwortlicher
Zeitraum: März 2026 – Heute

Projektaufgaben:

- Analyse der Risiken unkontrollierter agentischer Systeme (Shadow AI, fehlende Identifikation und Auditierbarkeit, Verletzung des Least-Privilege-Prinzips) und Ableitung eines risikobasierten Governance-Rahmens
- Konzeption eines eigenen Identitätstyps „Agent“ mit definiertem Identity-Lifecycle, dezentralem Ownership-Modell (Agent-, Capability- und Data-Owner) und risikotier-gesteuerter Governance-Intensität
- Entwurf einer Vier-Ebenen-Zielarchitektur (Control Plane im IAM, Decision Plane (PDP), Enforcement Plane, Kontrollen in den Zielsystemen) inkl. Policy as Code standard.
- Erhebung und Dokumentation eines Anforderungskatalogs mit technischen, organisatorischen, regulatorischen und datenschutzrechtlichen Anforderungen; Validierung von Zielbild und Governance-Anforderungen mit IAM-, dmGPT- und Analytics-Stakeholdern
- Konzeption eines Agent-Identity-MVP (governte Agenten-Identitäten aus dem IAM über Microsoft Entra Agent ID bis in die Zielsysteme) sowie eines MCP-Server-MVP mit konkretem Use Case
- Formulierung konkreter Handlungsempfehlungen und einer Roadmap als Grundlage für das Folgeprojekt zur Implementierung des Agentic IAM bei dm

Eingesetzte Technologien: One Identity, Microsoft Entra Agent ID, MCP (Model Context Protocol), dmGPT, Microsoft Office

Aufbau eines KI-Ökosystems und Agent-Gateways zur Befähigung interner AI-Teams bei der Umsetzung von GenAI- und agentischen Use Cases

Kunde: Mercedes-Benz Tech Innovation GmbH
Projektrolle: Lead Engineer (DevOps & AI), Solution Architect
Zeitraum: Januar 2025 – März 2026

Projektaufgaben:

- Technische Gesamtverantwortung für ein 13-köpfiges Team (7 Entwickler) inklusive Architekturentscheidungen, Priorisierung und Qualitätssicherung
- Design und Implementierung einer skalierbaren, serverlosen Backend-Architektur auf AWS mittels Infrastructure as Code (Terraform)
- Entwicklung und Betrieb von Microservices zur effizienten Verarbeitung und Orchestrierung von Frontend-Requests
- Integration und Anbindung interner Daten- und ML-Plattformen auf Azure, GCP und AWS

- Konzeption und Umsetzung eines dezentralen Agent-Gateways zur Integration und Kommunikation verteilter Agenten im Konzern
- Enge Zusammenarbeit mit cross-funktionalen Teams sowie technischen und fachlichen Stakeholdern

Eingesetzte Technologien: AWS SageMaker Unified Studio, AWS CDK, Terraform, AWS Lambda, API Gateway, DynamoDB, IAM, AWS Step Functions, Grafana, React, GCP Vertex AI, Cursor, Git, GitHub, Jira

Design und Implementierung eines skalierbaren, serverlosen Backends für eine GenAI-Mobile-App auf AWS

Projektname: Leap, Deutschland
Projektrolle: Co-Founder und Lead Engineer (DevOps & ML)
Zeitraum: August 2023 – Oktober 2023

Projektaufgaben:

- Design einer skalierbaren, vollständig serverlosen Backend-Architektur auf AWS
- (Lambda, API Gateway, DynamoDB, EventBridge Scheduler, SNS, WAF)
- Implementierung der Infrastruktur als Code (AWS CDK)
- Prompt Engineering und Prototyping von LLM-basierten Modellen
- Integration der OpenAI API sowie externer Services außerhalb der AWS-Umgebung
- Aufbau eines umfassenden Monitorings- und Alarming-Setups

Eingesetzte Technologien: Python, AWS CDK, Lambda, API Gateway, DynamoDB, EventBridge Scheduler, SNS, WAF, GitHub, DVC, Metaflow, Just, Webflow

Architektur und Aufbau einer skalierbaren End-to-End-MLOps-Plattform auf AWS

Projektname: q·ops
Projektrolle: Founder & CEO, Lead Engineer (MLOps & Cloud)
Zeitraum: August 2022 – August 2024

Projektaufgaben:

- Aufbau und Leitung eines eigenen Startups mit Fokus auf die Standardisierung und Beschleunigung von ML-Produktivsetzungen
- Entwicklung eines Businessplans, Erstellung von Anwenderberichten sowie systematische Sammlung und Priorisierung von KI-Use-Cases aus Kundenprojekten
- Entwurf, Vergleich und Prototyping verschiedener Cloud-nativer Architekturen für Model Deployment und Betrieb
- Implementierung einer skalierbaren, serverlosen Infrastruktur auf AWS mittels Infrastructure as Code
- Entwicklung einer wiederverwendbaren Repository-Vorlage inkl. eigenem Python-Package und Automatisierungen

- Umsetzung eines vollständigen End-to-End-MLOps-Stacks
  (CI/CD für Code und Modelle, Staging-Umgebungen, Paket- und Artefaktverwaltung, Monitoring und Alerting)
- Konzeption und Aufzeichnung eines Onlinekurses zur Schulung von Anwendern in q·ops und AWS-basierter MLOps-Architektur

Eingesetzte Technologien: Python, AWS CDK, AWS SageMaker, Lambda, ECS, AWS Batch, Step Functions, API Gateway, DynamoDB, IAM, VPC, CloudWatch, GitHub, DVC, Metaflow, Just, Webflow

Bewertung von Anbietern und Architekturentwurf einer hochskalierbaren IoT-Datenplattform zur Überwachung von über 3 Mio. Containern

Kunde: Hapag-Lloyd, Deutschland
Projektrolle: Senior Data Scientist, Project Lead
Zeitraum: April 2022

Projektaufgaben:

- Erhebung und Strukturierung von User Stories und Business Use Cases aus mehreren Fachbereichen
- Definition von Plattform-Capabilities und Anforderungen als Bewertungsraster für potenzielle IoT-Lösungen
- Durchführung von Interviews mit Stakeholdern und Entwicklerteams zur Identifikation zentraler Herausforderungen bestehender und geplanter IoT-Projekte
- Bewertung von Make-, Buy- und Hybrid-Ansätzen (On-Premises, vollständige Vendor-Lösungen, Cloud Platform as a Service)
- Diskussion und Ableitung geeigneter Technologie-Stacks sowie konkrete Tooling-Empfehlungen
- Konzeptionelle Erstellung von Cloud-Architekturdiagrammen auf AWS und GCP in enger Zusammenarbeit mit Cloud Engineers
- Priorisierung der Maßnahmen und Erstellung einer umsetzungsnahen Entwicklungs-Roadmap

Eingesetzte Technologien: AWS, GCP, Microsoft Office, Miro, draw.io

Reifegradbeurteilung und Konzeption einer AWS-basierten KI-Plattform in enger Zusammenarbeit mit dem KI-Center of Excellence

Kunde: Hapag-Lloyd, Deutschland
Projektrolle: Senior Data Scientist
Zeitraum: Januar 2022 – März 2022

Projektaufgaben:

- Erhebung und Strukturierung von Plattform-Capabilities und Anforderungen auf Basis bestehender KI-Use-Cases sowie Erfahrungen aus früheren Projekten
- Systematische Reifegradbeurteilung vorhandener KI-Plattform- und Deployment-Lösungen anhand eines detaillierten Bewertungsrasters
- Identifikation zentraler technischer und organisatorischer Hürden bei der Produktivsetzung von KI-Use-Cases in Entwicklerteams

- Diskussion und Bewertung geeigneter Technologie-Stacks sowie Ableitung konkreter Tooling-Empfehlungen
- Konzeptionelle Erstellung von AWS-Architekturdiagrammen für unterschiedliche Ausbaustufen (Light- und Full-Scale-Model-Serving, -Updating und -Retraining)
- Priorisierung der Maßnahmen und Erstellung einer umsetzungsnahen Entwicklungs-Roadmap

Eingesetzte Technologien: AWS, Microsoft Office, draw.io

Predictive-Maintenance-PoC zur Erkennung von Radschäden mittels Gleissensorik und SAP-Schadcodes

Kunde: Deutsche Bahn, Deutschland
Projektrolle: Senior Data Scientist
Zeitraum: Oktober 2021 – Januar 2022

Projektaufgaben:

- Durchführung fachlicher Interviews mit Experten und Ingenieuren zur Domänen- und Anforderungsanalyse
- Aufbereitung, Synchronisation und Zusammenführung hochfrequenter Zeitreihendaten aus Drucksensoren mit Schadcode-Labels aus SAP
- Deskriptive Analyse, Datenbereinigung und Qualitätssicherung der Mess- und Labeldaten
- Feature Engineering und Entwicklung eines Baseline-Modells auf Basis von Random Forest
- Training eines Autoencoder-Modells zur Reduktion der dimensionalen Komplexität um ca. 64 %
- Entwicklung eines robusten „Rolling-RANSAC“-Ansatzes zur zeitlichen Vorhersage von Radschäden mit erhöhter Ausreißerresistenz
- Strukturierung, Bereinigung und Dokumentation des Projekt- und Team-Repositories

Eingesetzte Technologien: AWS, Python, Pandas, Scikit-learn, PyTorch, VS-Code, Git, GitLab

Planung und Entwicklung einer KI-Plattform mit Model-Serving-Blueprint zur Standardisierung von ML-Deployments

Kunde: Vodafone, Deutschland
Projektrolle: Senior Data Scientist & Team Lead
Zeitraum: Mai – August 2021

Projektaufgaben:

- Durchführung von Research sowie Interviews zu MLOps-Prozessen zur Identifikation wiederkehrender Aufwände und Anforderungen
- Konzeption einer standardisierten KI-Plattform unter Nutzung von Open-Source-Komponenten und AWS-cloud-nativen Services
- Prototypische Evaluation verschiedener Model-Serving- und automatisierter Retraining-Ansätze
- Entwicklung und Iteration eines AWS-Ressourcen-Blueprints als automatisiertes Fundament für zukünftige ML-Model-Deployments

- Technische Validierung der durch Cloud Engineers implementierten AWS-Architektur inkl. Service-Berechtigungen und Sicherheitskonzepten
- Konzeption und Verprobung eines Data-Science-Model-Serving-Blueprints (u. a. mit Apache Airflow, AWS SageMaker und DVC) als Leitfaden für zukünftige Teams
- Erstellung einer detaillierten technischen Dokumentation und Benutzerhandbücher zur Nutzung der KI-Plattform
- Aufbereitung der Ergebnisse und Architekturentscheidungen in Form von Management- und Stakeholder-Präsentationen

Eingesetzte Technologien: AWS SageMaker, S3, API Gateway, Lambda, RDS, Python, Apache Airflow, DVC, Git, GitHub, VS Code

Proof of Concept einer Web-App zur Informationsextraktion (NER) aus Dokumenten mit Human-in-the-Loop-Ansatz

Kunde: Vodafone, Deutschland
Projektrolle: Senior Data Scientist
Zeitraum: März 2021

Projektaufgaben:

- Erhebung und Strukturierung funktionaler Anforderungen für eine NER-basierte Informationsextraktion
- Konzeption eines Human-in-the-Loop-Ansatzes zur iterativen Verbesserung der Extraktionsergebnisse
- Implementierung einer Dash-Wrapper-Komponente zur Integration in eine React-basierte Frontend-Umgebung
- Umsetzung einer Dash-basierten Web-Applikation mit API-Schnittstelle zum Backend

Eingesetzte Technologien: Python, Dash, FastAPI, Swagger (OpenAPI), VS Code, Git, GitHub

Auslegung und Planung einer datengetriebenen, KI-gestützten Applikation zur Wirkstoffentwicklung

Kunde: Merck Group, USA
Projektrolle: Senior Data Scientist
Zeitraum: Januar – Februar 2021

Projektaufgaben:

- Planung und Durchführung strukturierter Interviews und Workshops mit Chemikern, Stakeholdern und potenziellen Anwendern zur Anforderungserhebung
- Übersetzung fachlicher und nutzerbezogener Anforderungen in User Stories und User Journeys
- Analyse bestehender marktüblicher Lösungen und Referenzanwendungen als Grundlage für das eigene UX-Konzept
- Konzeption und Erstellung eines Applikations-Mockups in enger Zusammenarbeit mit UX/UI-Designern

- High-Level-Design der technischen Gesamtlösung in Abstimmung mit DevOps-Experten inkl. initialer Abschätzung von Komplexität und Umsetzungsaufwand
- Erstellung eines Request-for-Proposal-Dokuments (RFP) zur Vorbereitung der Implementierungsphase

Eingesetzte Technologien: Figma, Microsoft Office

Bewertung des Potenzials zur Rekonstruktion von Trainingsdaten durch Angriffe auf ML-Modelle in der Arzneimittelentwicklung

Kunde: Merck Group, USA
Projektrolle: Data Scientist
Zeitraum: Oktober – Dezember 2020

Projektaufgaben:

- Analyse und Bewertung von Risiken des Reverse Engineerings von Trainingsdaten in ML-Modellen zur Arzneimittelentwicklung
- Literaturrecherche und systematische Einordnung bekannter Angriffsmechanismen (Membership Inference, Model Extraction, Model Inversion, Property Inference, Molecular Reconstruction)
- Bewertung von Angriffsvoraussetzungen, Relevanz und potenzieller Schadenshöhe im konkreten Kundenkontext
  (Datenbasis, Modellzugriff, Bedrohungsszenarien)
- Durchführung ausgewählter, hochrelevanter Angriffe unter Verwendung öffentlich verfügbarer bioaktiver Moleküldatenbanken
- Ableitung technischer und kommerzieller Maßnahmen zur Risikominderung und Absicherung der Modellintegrität
- Entwicklung einer Python-Bibliothek zur systematischen Bewertung der Sicherheit individueller ML-Modelle
- Implementierung geeigneter Schutzmaßnahmen in der Kundensoftware

Eingesetzte Technologien: Python, Pandas, Scikit-learn, VS Code, Git, KVM

Deployment eines BERT-basierten NLP-Modells zur E-Mail-Klassifikation im „Dispute Capturing“-Prozess der Containerschifffahrt

Kunde: Hapag-Lloyd, Deutschland
Projektrolle: Data Scientist
Zeitraum: Mai – Juli 2020

Projektaufgaben:

- Weiterentwicklung und Erweiterung der im vorherigen Projekt etablierten Cloud-Architektur auf AWS
- Aufwandsschätzung, technische Projektplanung und Priorisierung
- Implementierung von Abstraktions- und Adapter-Komponenten zur Reduktion AWS-spezifischer Abhängigkeiten

- Weiterentwicklung der Datenaufbereitung sowie Training und Feinabstimmung eines BERT-basierten Modells bis zur Produktionsreife
- Implementierung zustandsloser, skalierbarer Container für Inferenz und Vorverarbeitung
- Review und Integration aller relevanten AWS-Komponenten
- (Lambda, API Gateway, ECS, CodeBuild CI/CD Pipeline, CloudWatch, DynamoDB, S3)
- Durchführung von End-to-End-Tests und Unterstützung der ProduktivsetzungEingesetzte Technologien: AWS Cloud Services, Python, Pandas, Transformers (BERT), VS-Code, Git

Eingesetzte Technologien: AWS Cloud Services, Python, Pandas, Transformers (BERT), VS Code, Git, GitHub

Deployment eines NLP-Modells zur E-Mail-Klassifikation im Buchungsprozess der Containerschifffahrt

Kunde: Hapag-Lloyd, Deutschland
Projektrolle: Data Scientist
Zeitraum: Januar – März 2020

Projektaufgaben:

- Konzeption einer skalierbaren Cloud-Architektur auf AWS für den produktiven Einsatz eines NLP-Modells
- Aufwandsschätzung, technische Projektplanung und fachliche Priorisierung
- Fachliche Leitung des Entwicklerteams mit Product-Owner-Verantwortung
- Weiterentwicklung von Datenaufbereitung und Random-Forest-Modell bis zur Produktionsreife
- Implementierung zustandsloser, skalierbarer Container für Datenbereinigung und Inferenz
- Review und Integration aller relevanten AWS-Komponenten (Lambda, API Gateway, ECS, CodeBuild CI/CD Pipeline, CloudWatch, DynamoDB, S3)
- Durchführung von End-to-End-Tests und Unterstützung der Produktivsetzung
- Aufbereitung und Präsentation der Ergebnisse gegenüber dem Kunden sowie Vorträge auf dem PyData und NLP Meetup Hamburg (Januar 2021)

Eingesetzte Technologien: AWS Cloud Services, Python, Pandas, Scikit-learn (Random Forest), Git, GitHub

Rapid Prototyping und Business Case Analyse für 6 NLP Use Cases

Kunde: Hapag-Lloyd, Deutschland
Projektrolle: Data Scientist
Zeitraum: Juli – September 2019

Projektaufgaben:

- Analyse und Priorisierung von sechs NLP-Use Cases in Zusammenarbeit mit Fachabteilungen
- Durchführung strukturierter Interviews mit Subject Matter Experts zur fachlichen und datengetriebenen Use-Case-Definition
- Integration, Bereinigung und Harmonisierung heterogener Datensätze aus sechs unterschiedlichen Unternehmensbereichen

- Datenaufbereitung und Feature Engineering für NLP-Modelle
- Entwicklung mehrerer Prototypen (PoCs) inkl. Modelltraining und -evaluation
- Auswahl geeigneter Modellansätze abhängig von Business-Case-Zielen und Zielvariablen
- Iterative Validierung, Ergebnisbewertung und Präsentation der Resultate gegenüber Stakeholdern
- Umsetzung vollständiger Prototypenzyklen (Datenaufbereitung → Modelltraining → Evaluation) innerhalb von jeweils ca. 2–3 Wochen

Eingesetzte Technologien: AWS SageMaker, Python, Pandas, Scikit-learn, Jupyter Notebooks

PoC Anomalie Erkennung in Sensordaten von Bahninfrastruktur

Kunde: Vossloh, Deutschland / Frankreich
Projektrolle: Data Scientist
Zeitraum: April – Mai 2019

Projektaufgaben:

- Aufbereitung, Konvertierung und Analyse hochfrequenter Sensordaten
- Entwicklung eines PoC zur Anomalieerkennung mittels unsupervised ML-Verfahren
- Identifikation und Bewertung potenziell kritischer Ereignisse in der Infrastruktur
- Visualisierung der Ergebnisse zur fachlichen Einordnung und Entscheidungsunterstützung
- Erstellung von Visualisierungen

Eingesetzte Technologien: Python, Pandas, Scikit-learn, Jupyter

Topic Clustering und Identifikation geschäftsrelevanter Inhalte

Kunde: Hapag-Lloyd, Deutschland
Projektrolle: Data Scientist
Zeitraum: Feb – März 2019

Projektaufgaben:

- Aufbereitung und Analyse unstrukturierter Textdaten aus E-Mails und JIRA-Tickets
- Entwicklung von Topic-Clustering-Ansätzen zur thematischen Strukturierung großer Textmengen
- Klassifikation und Filterung von E-Mails nach Geschäftsrelevanz
- Explorative Analyse zur Ableitung von Insights und Entscheidungsgrundlagen
- Entwicklung und Validierung datengetriebener Business Cases
- Entwicklung und Validierung von Business Cases

Eingesetzte Technologien: Python, Scikit-learn, Jupyter Notebooks