

## Persönliche Daten

✉ sks.sauravs@yahoo.com

🌐 sksdotsauravs

🏠 WH 762-11-01-17

Siegmunds Hof 2-4

10555 Berlin

Deutschland

♂ Männlich

## Kenntnisse und Fähigkeiten

Python

R

Java

HTML

CSS

NumPy

Pandas

PyTorch

HuggingFace

Flair

Matplotlib

Plotly

Spring Framework

Apache Maven

MySQL

Redis

Hibernate

JUnit

Mockito

Git

Jenkins

Rundeck

Docker

Kubernetes

Grafana

# Saurav Kumar Saha

## Berufserfahrung

### Studentische Hilfskraft - NLP und Datenschutzforschung

#### Cognitive Algorithms Lab, BHT 🌐

📅 Dezember 2024 – März 2025

📍 Berlin, Deutschland

- Entwurf und Implementierung eines **einheitlichen Pseudonymisierungstools** für deutsche Texte unter Verwendung des **mT5**-Sprachmodells
- Bewertung der Modellleistung durch **Benchmarking** mit klassischen **NER**-Modellen und **LLM Prompt-Engineering**-Techniken für 10K+ deutsche E-Mail-Texte
- Einsatz einer leichtgewichtigen Textpseudonymisierungs-App mit **REST API** (FastAPI) und **Webinterface** (Streamlit) auf einem universitären **Kubernetes**-Cluster, die die Generierung von 5 pseudonymisierten Varianten pro Benutzereingabe ermöglicht
- Mitautor eines **Forschungspapiers**, das auf der ersten internationalen Konferenz über KI in Medizin und Gesundheitswesen in Medicine and Healthcare (AiMH 2025) angenommen wurde, das sich mit dem **Datenschutz** im klinischen NLP beschäftigt,   
🔗 [GitHub](#)

### Werkstudent - Softwareentwicklung

#### Solactive AG 🌐

📅 Februar 2020 – August 2023

📍 Berlin, Deutschland

- Entwicklung einer automatisierten Pipeline für die Erstellung von Financial **Factsheets** mit Python, Docker und Rundeck - Reduktion der Lieferzeiten um 90% von Stunden auf Minuten
- Einführung einer lokalen Entwicklungsumgebung mit **Docker**, um die Einarbeitungszeit für neue Entwickler um 70% zu reduzieren und manuelle Einrichtungsschritte zu vermeiden
- Entwicklung von mehr als 5 **Jenkins**-Testautomatisierungspipelines, wodurch der manuelle Testaufwand in anderen Teams um etwa 60% reduziert wurde, was sich direkt auf die Software-Release-Zyklen auswirkte
- Implementierung einer Funktion zum Zurücksetzen der Zeit für die Plattform Test Systems, einer auf **Microservices** basierenden Java-Anwendung, die mit **Spring Boot** und **MySQL** erstellt wurde und das Zurücksetzen von mehr als 100 täglichen Testumgebungen ohne manuellen Eingriff ermöglicht
- Erstellung von **Grafana**-Dashboards mit **Flux** zur Überwachung der Leistung und des Zustands von über 20 wichtigen Backend-Diensten

### Senior Software Engineer - Backend

#### Innovative Solutions 🌐

📅 August 2018 – August 2019

📍 Dhaka, Bangladesch

- Leitung der Entwicklung des **Freeway**-Projekts: Online-**Ticket-Reservierungs- und -Kaufsystem** für Busse mit mehr als 30 verschiedenen Zielen von Shohagh Paribahan Limited, einem führenden Busdienst in Bangladesch
- Verwalten der Entwicklung von **Spring Boot + MySQL**-basierten Backend-Diensten und APIs, die den gesamten Lebenszyklus abdecken: Anforderungsanalyse, Design, Implementierung, Testen, Refactoring und Bereitstellung in Entwicklungs- und Produktionsumgebungen mit **Docker** und **TeamCity**

## Sprachen

Deutsch	● ● ● ● ●
Englisch	● ● ● ● ●
Bengalisch	● ● ● ● ●

## Software Engineer

**BizMotion Limited** 

📅 Februar 2013 – Mai 2018

📍 Dhaka, Bangladesch

- Konzeption und Leitung der Entwicklung des **SMC-Program** projekts, mit dem über 500 Pharmareferenten Umfragen, Bestandsaktualisierungen und Schulungsmaterialien über ein **Spring Boot** + **PostgreSQL** Backend verwalten können
- Entwicklung von Backend-Diensten und APIs für das Projekt **BizPharma**, eine Lösung zur Automatisierung des Außendienstes, die von mehr als 20 **FMCG** Unternehmen zur Rationalisierung der Arbeitsabläufe in den Bereichen Lagerhaltung, Verkauf und Vertrieb genutzt wird
- Full-Stack-Entwickler des VoIP Dialer CMS-Projekts - **Spring-Framework** und **PostgreSQL**-basierte Webanwendung für rollenbasiertes Zugangs- und Betriebsmanagement von VoIP-Ressourcen

## Ausbildung

M.Sc. in Data Science

**Berliner Hochschule für Technik**

📅 Oktober 2019 – März 2025

📍 Berlin, Deutschland

B.Sc. in Computer Science and Engineering

**Shahjalal University of Science and Technology**

📅 Januar 2008 – Dezember 2011

📍 Sylhet, Bangladesch

## Veröffentlichungen

- [1] Saurav Kumar Saha und Felix Biessmann. „End-to-end Pseudonymization of German Texts with Deep Learning – An Empirical Comparison of Classical and Modern Approaches“. In: *Proceedings of the 1st International Conference on AI in Medicine and Healthcare (AiMH' 2025)*. Hrsg. von Sergey Y. Yurish. Innsbruck, Austria, 8–10 April 2025. Barcelona, Spain: IFSA Publishing, S. L., 2025, S. 98–104. ISBN: 978-84-09-71190-1. DOI: 10.13140/RG.2.2.20387.77606.