Simon Kutsche

LinkedIn | (954) 802-9906 | simonkutsche.com | simon.kutsche@gmail.com | GitHub

Skills

- Python | SQL | Linux | PyTorch | TensorFlow | Keras | PySpark | Flask | NumPy | SciPy | scikit-learn | Pandas | Matplotlib | Tableau | Git
- OpenCV | Hadoop | Spark | CI/CD | OpenMPI | NoSQL | Docker | Kubernetes | PyCharm | VS Code | JavaScript | HTML | CSS
- AWS Elastic Beanstalk, Codepipeline, Lambda, Sagemaker | Databricks | English, German - *Native proficiency* | Spanish - *Conversational*

Experience

Machine Learning Engineer **Robert Bosch LLC** *Hildesheim, Germany* **01/2023 - 09/2023**

• Developed a Deep Learning-based process monitoring system using LSTM-Autoencoder for real-time anomaly detection in multivariate sensor data, contributing to increased availability and prevention of damage to expensive workpieces.
• Extended the LSTM-Autoencoder model's capabilities by training it on datasets from multiple identical CNC machines, achieving a remarkable average F1-score of 0.96 across monitoring scenarios.
• Utilized AWS SageMaker for model training and deployment, integrated AWS IoT for data ingestion, and implemented CI/CD pipelines using AWS CodePipeline to ensure continuous improvement and deployment efficiency.

Deep Learning Engineer **Robert Bosch LLC** *Renningen, Germany* **03/2022 - 09/2022**

• Led the conception and development of cutting-edge loss weighting methods for multi-task learning in autonomous driving, showcasing superior performance compared to industry benchmarks.
• Conducted extensive evaluations on scene understanding tasks, including semantic segmentation, object detection, depth estimation, and surface normal estimation. Leveraged datasets like Cityscapes and NYUv2 to establish a robust evaluation framework.
• Played a crucial role in integrating novel loss weighting methods seamlessly into the existing code base of autonomous driving software. Demonstrated a deep understanding of software architecture to ensure compatibility and performance optimization.
• Documented outcomes in a comprehensive paper submitted to CVPR 2023, contributing to academic advancements in the intersection of neural networks and autonomous driving. Recognized for innovation with two registered patents, solidifying the intellectual property value of implemented methods.

Data Engineer **Scalefree International GmbH** *Hanover, Germany* **10/2020 - 01/2022**

• Led the internal Business Intelligence development team in the capacity of a Scrum Master, orchestrating efficient and collaborative processes for the successful execution of Business Intelligence initiatives.
• Spearheaded the development of streamlined processes for loading the staging area and the raw data vault, utilizing AWS services such as S3, Lambda, and Batch. Achieved enhanced efficiency and reliability in data workflows.
• Employed Tableau in a collaborative data visualization project aimed at analyzing customer engagement and optimizing marketing strategies for a leading e-commerce platform (external customer), resulting in a 15% increase in conversion rates within six months.

Machine Learning Engineer Intern **Novatec Consulting GmbH** *Hanover, Germany* **11/2019 - 06/2020**

• Spearheaded the development of a churn prediction prototype application, showcasing proficiency in utilizing various machine learning algorithms, including Random Forest, SVM, Gradient Boosted Decision Trees, and Logistic Regression.
• Conducted a comprehensive evaluation of machine learning algorithms to identify optimal solutions for churn prediction. Analyzed and compared the performance of algorithms to enhance predictive accuracy.
• Led the assessment of Cloud Machine Learning Services, comparing MS Azure Databricks with PySpark, AWS Sagemaker, and Google Cloud BigQuery. Evaluated their suitability for the specific requirements of the churn prediction application.

Education

Bachelor of Science **University of Hanover** *Hannover, Germany* **03/2016 - 06/2020**

- Major in Business Information Systems, minor in Business Intelligence (GPA: 3.5)

Master of Science **University of Hildesheim** *Hildesheim, Germany* **04/2020 - 01/2023**

- Major in Data Analytics, minor in Advanced Computer Vision (GPA: 3.5)

Lecturing

• Delivering comprehensive lectures and hands-on sessions in the Data Science Bootcamp of 4GeeksAcademy at the Miami-Dade College, imparting practical knowledge in machine learning, data analysis, and statistical modeling. Mentoring students through real-world projects, fostering their skills and understanding in the rapidly evolving field of data science. **(11/2023 - now)**

Simon Kutsche

LinkedIn | (954) 802-9906 | simonkutsche.com | simon.kutsche@gmail.com | GitHub

Academic Projects

- ***Generative Adversarial Networks***: Conducted image-to-image translation between the domains of regular images and artworks with Deep Generative Adversarial Networks. Enhanced CycleGAN by introducing a two-objective discriminator as regularization, incorporating an adversarial self-defense for better cycle-consistency, and applying differentiable augmentation on the target domain with less data. Employed agile intercultural project management techniques to manage the project successfully. Designed and developed a Flask-based web application to deploy and showcase the model, enabling users to generate art pieces based on their input images, enhancing user engagement and interactivity **(12/2020 - 03/2022)**

Coursework

- ***Natural Language Processing***: Developed a sentiment analysis model to automatically categorize customer reviews into positive, negative, or neutral sentiments. Generated a dataset through efficient labeling with Azure Machine Learning Data Labeling and storage using Azure Blob Storage. Conducted data preprocessing on Azure Databricks for text cleaning and feature extraction and utilized Azure NLP services for sentiment analysis model selection and fine-tuning. **(10/2021 - 11/2022)**
- ***Machine Learning***: Implemented various machine learning models such as ridge regression with SGD, LASSO with coordinate descent, least-angle regression, logistic regression with newton method, gradient boosted decision tree, and AdaBoost from scratch in Python and NumPy on real-world datasets like Rossmann sales and Wine quality data. Employed data preprocessing techniques such as one-hot encoding, stratified sampling, PCA, and KNN data imputation. Performed exploratory data analysis on various real-world datasets using Pandas and Matplotlib. **(04/2020 - 09/2021)**
- ***Deep Learning/Computer Vision***: Trained a CNN end-to-end on a self-driving dataset (camera view from the car) using regularization techniques such as cutout and mixup, and implemented a custom batch normalization layer and residual connections to predict the steering angle in PyTorch. Compared metric learning techniques such as learned embedding of a simple classification model, contrastive loss, and triplet loss with an embedding layer for MNIST data using TensorFlow. Implemented transfer learning for training a U-Net model on a real-world weed field image dataset with a custom categorical cross-entropy loss. Pretrained the first half of the model on the classification dataset DeepWeeds using TensorFlow, improving the test accuracy by 1.5\% compared to a vanilla U-net model and visualized the predicted segmentation map. Generated adversarial examples using the Carlini-Wagner attack against a CNN trained on MNIST data and created sparse perturbations with the Hoyer-Square regularizer using PyTorch. Implemented YoloV1 from scratch for object detection on a real-world urban street scenes dataset to detect pedestrians **(04/2020 - 09/2021)**
- ***Distributed Computing***: Performed exploratory data analysis using PySpark on the movielens 10m dataset and used the Hadoop MapReduce framework on BTS flight data. Conducted distributed K-means clustering and distributed linear regression using SGD on KDD Cup 1998 dataset and VirusShare executables with OpenMPI, including a performance analysis on the speed-up with different numbers of used cores. Implemented Naive Bayes and SVM classifiers from scratch to categorize news items on 20 newsgroups text dataset using preprocessing techniques such as bag-of-words and TF-IDF feature representation and the Hadoop MapReduce framework. Employed distributed matrix factorization using coordinate descent with the Hadoop MapReduce framework on the movielens 10m dataset. **(04/2020 - 03/2021)**
- ***Reinforcement Learning***: Utilized PyTorch to develop both the Deep Q-Learning model and the REINFORCE algorithm with policy gradients from scratch to solve the Gym environment Mountain Car. **(10/2022 - 03/2023)**