Saikat Mitra

Stuttgart, Germany | +49 17659691594 | saikatmitra1998@gmail.com | linkedin.com/in/saikat-mitra |
https://github.com/saikatmitra1998

Education

Otto von Guericke University — Magdeburg, Germany
Master of Science in Data and Knowledge Engineering (GPA 1.8) — *April 2022 - Present*

St. Xavier's College, Kolkata | University of Calcutta — Kolkata, India
Bachelor of Science in Computer Science (GPA 2.1) — *June 2016 - June 2019*

Experience

Deep Learning | Master Thesis — December 2023 - Present
German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt e.V.) — *Stuttgart, Germany*

- Utilizing 'Mobility in Deutschland' and 'German Mobility Panel' data to construct an open-licensable dataset using deep learning techniques.
- Focused on analyzing energy consumption patterns of electric vehicles in the daily lives of individuals in Germany.
- Conducting data preprocessing, including cleaning and transformation, using Pandas, NumPy, and Dask libraries.
- Contributed enhancements to the open-source tool Vencopy, employing object-oriented programming (OOP) principles in Python to generate time series data from mobility patterns.
- Investigating patterns and trends by implementing techniques such as LSTM, Transformers and XGBoost on time series data using PyTorch framework.

Software Developer | Working Student — Jan 2023 - Present
Otto von Guericke University — *Magdeburg, Germany*

- Integral member - Bildungsraum Digital (BIRD) project by Federal Ministry of Education and Research (BMBF).
- Developed and optimized ABAP programs and interfaces to enhance SAP functionality and business requirement, following Agile development practices.
- Created OData services facilitating seamless data exchange between SAP system and external application.
- Designed and delivered responsive web application using SAP UI5 framework, enhancing user experience.
- Utilized Docker to containerize environments, ensuring consistency across development, testing, and deployment.

Graduate Teaching Assistant | Data Mining — April 2023 - July 2023
Otto von Guericke University — *Magdeburg, Germany*

- Assisted Prof. Myra Spiliopoulou (KMD dept.) in teaching the course on Data Mining to graduate students.
- Created engaging exercises, problem sets, and supplementary materials to boost student comprehension.
- Conducted exercise classes for Masters student teaching them different Machine Learning algorithms.
- Contributed to refining the curriculum to align with industry trends, ensuring a pertinent learning experience.

Software Engineer | Systems Engineer | CBO — Aug 2019 - March 2022
Deloitte US India Consulting Pvt. Ltd. — *Mumbai, India*

- Utilized Python scripting for automation tasks, including server monitoring, log analysis, and data manipulation.
- Used Shell Script to automate email alerts for server downtime, ensuring rapid stakeholder notification.
- Efficiently designed and implemented robust SQL scripts for use in production.
- Employed Bamboo and Jenkins for Continuous Integration, expediting testing and software delivery.
- Automated QA and production deployment using Ansible playbooks and Python scripts.
- Used Docker Compose, orchestrated multi-container applications and simplifying microservices deployment.

Technical Skills

Programming Languages: Python, Java, R, Shell Scripting, Oracle SQL, MySQL, JavaScript, HTML/CSS, Ansible
Data Science: Machine Learning, Deep Learning, Natural Language Processing, Recommender System
Libraries: TensorFlow, Keras, PyTorch, Scikit-learn, OpenCV, spaCy, NLTK, Captum, Pandas, NumPy, Matplotlib, MLflow, Wandb
Frameworks: Node.js, Flask, FastAPI, Streamlit
Developer Tools: Git, SVN, Docker, Google Cloud Platform, Azure, Jenkins, Bamboo, PyCharm, IntelliJ, Eclipse, Power BI, SonarQube, JIRA, Linux, Miro
Azure: Data Factory, Databricks, Data Explorer, HDInsight, Azure DevOps
Google Cloud Platform: App Engine, Cloud Run, VertexAI, AutoML, BigQueryML, Data studio(Looker)
AWS: SageMaker, S3, EC2
Microsoft: Excel, Outlook, SharePoint, Powershell

Projects

Validating Attribution Techniques using Deep Learning | *CNN, XAI, PyTorch* May 2023 – Sept 2023

- Aims to validate the robustness of saliency attribution techniques, providing explanations for model predictions.
- Randomly sampled 113 images from the training set and generated over 16,000 perturbed images, providing a robust foundation for conducting experiments and assessing model performance on unseen data.
- Implemented various attribution methods including GradCAM, SmoothgradCAM++, ScoreCAM and LayerCAM and tested it against ResNet18, ResNet50, VGG16, InceptionV3**.**
- Used occlusion-based perturbation and adversarial noise assessment to reveal vulnerabilities and limitations in popular attribution methods, clarifying model interpretability, region sensitivity, and saliency map reliability.
- Leveraged advanced metrics including Pearson Correlation Coefficient, Earth Mover's Distance, Area under Curve, Similarity, and Normalised Scanpath Saliency to provide comprehensive insights and rigorous evaluation.

Spam Detection using LSTM (Long Short-Term Memory) | *RNN, TensorFlow* Jan 2024

- Conducted data preprocessing, including cleaning and transformation, utilizing the Pandas library.
- Implemented LSTM architecture followed by a dense layer employing a sigmoid activation function.
- Utilized the Adam optimizer, binary crossentropy loss function, and accuracy as the evaluation metric.
- Visualized the loss and accuracy metrics using Matplotlib for both training and validation datasets.

COVID-19 Data Analytics and Reporting using Azure | *Data Factory, Databricks, PowerBI, SQL* Dec 2023

- Built a solution architecture using Azure Data Factory, Azure Data Lake, Databricks, PoweBI and HDInsight.
- Integrated data from HTTP clients, Azure Blob Storage and Azure Data Lake Gen2 using Azure Data Factory.
- Created ADF pipelines to execute HDInsight activities and Databricks Notebook to carry out data transformations
- Created CI/CD pipeline for releasing Azure Data factory artefacts for deployment.

Chatbot using Transformers in PyTorch | *Transformers, PyTorch* Nov 2024

- Aimed to develop a QA chatbot leveraging Transformers architecture in PyTorch.
- Preprocessed conversational datasets to train chatbot, ensuring data quality and relevance for effective learning.
- Evaluated the transformer model's response generation capability through Greedy Decoding with batch size of 1, showcasing its effectiveness.

Employee Performance Optimisation using ML in GCP | *BigQuery, BigQueryML* May 2023 - July 2023

- Leveraged GCP tools to visualize employee learning curves and develop data-driven employee assignment.
- Demonstrated the benefits of data-driven employee assignment & improved allocation of service employees.
- Utilized BigQuery, AutoML, BigQuery ML and Vertex AI for data exploration and training, Google data Studio (Looker) for visualizations.
- Used Python and Flask to create the User Interface (UI) and deployed it in Google App Engine (GAE).
- Employed Google Distance Matrix API for calculation of distances to client sites.

Movie Recommendation System using ML | *Python, TKinter, SVD* Feb 2019 - June 2019

- Spearheaded team of three, building a Collaborative based Movie Recommendation System leveraging Python.
- Implemented the SVD algorithm to mitigate MSE and RMSE and enhance system accuracy.
- Utilized Tkinter, as GUI, to design and develop an intuitive and visually compelling user interface.

Certifications

Microsoft Certified Azure Fundamentals (AZ-900)

HackerRank Certification SQL

HackerRank Certification Python

Language Skills

English: Fluent (C1)
German: Beginner (A1)