Sourabh **Raj**

Data Scientist | Machine Learning Engineer

About me

A keen learner and an explorer willing to push boundaries to seek knowledge. I facilitate a working experience of machine learning concepts (Supervised/Unsupervised/Reinforcement learning) along with a demonstrated history in Full-stack development. With my problem-solving attitude and ability to learn and adapt quickly, I look forward to crafting cutting-edge AI solutions to cater to real-world needs.

Contact

sourabhraj1006@gmail.com

+49 1631686684

Motzstr. 27
10777 Berlin, Germany

sraj1006

rajs1006

Professional Skills

Machine Learning | Data Science | Deep Learning | Computer Vision | Semantic Segmentation | Image Translation | NLP | Python | Java | Database/SQL | MLOps | AWS | Airflow | Docker | Git CI/CD | Javascript/ReactJS | Unity3D/C# | C/C++ | Full Stack | Web Development

Soft Skills and Strengths

Quick Learner | Problem Solving | Hard Working | Ability to Plan and Organize | Leadership | Team Work | Autonomy | Adaptability

Languages

English - *Fluent*

German - *B1*

Hindi - *Native Language*

WORK EXPERIENCE

Apr 2020 - present

Machine Learning Specialist(Head of DSc) — Berlin

Resonanz energy GmbH

Plan, design, and prototype production ready machine learning solutions to support automation in the field of Renewable Energy like, real-time energy time series prediction to plan the energy production, decision-making algorithms to support energy trading.

- Python, Scikit-learn, Pytorch, Tensorflow, numpy, Pandas, Dask, Numba, PySpark, Flask
- AWS, Airflow, Pyspark, Dagster, TimescaleDB, Docker, kubernetes, Gitlab

Jul 2019 - Mar 2020

Machine Learning Engineer — Berlin

Funke Mediengruppe GmbH Co. KGaA

Design and develop ML projects for Churn Rate prediction, NLP-based Article Tagging, and Classification.

- Python, Pytorch, Keras, Tensorflow, Pandas, Numpy, Scikit-learn, statsmodel, Tensorboard, Plotly, Streamlit/Bokeh/Dash

Develop and deploy end-to-end ML models(Classification, Prediction, Deep neural network, Natural language processing) on cloud(GCP/AWS).

May 2019 - Jul 2019

Trainee Data Scientist — Berlin

Frequenz Energy-as-a-Service GmbH

Research and develop classification/prediction models for multivariate time-series data.

- Deep Neural Network, SVM, Adaboost, XGBoost, Random Forest, PCA, tSNE, Matplotlib

Mar 2018 - May 2019

Developer — Berlin

FUNKE Digital TV Guide GmbH

Develop and Optimize Java based Microservices and test framework.

- JAVA 8, Spring Boot Microservices, Multithreading, Docker, Kubernetes, Linux

Feb 2015 - Nov 2017

Sr. Technical Consultant — Gurugram

Deloitte USI

Design, Develop and Deploy web applications

- JAVA and JavaScript-based frameworks along with SQL queries.

Nov 2011 - Feb 2015

Web Developer — Gurugram

Publicis.Sapient

Research and work on POC for various Java-based frameworks

- JAVA, Spring, Spring Security, Hibernate, SQL queries, Shell Scripts.

EDUCATION

Dec 2017 - Feb 2020

Master of Science — Berlin, Germany

Technische Universität Berlin

Computer Science (Cognitive Systems)

Machine Learning(Supervised/Unsupervised Learning), Advanced Machine Learning - Theory and Application, Machine Learning Project - Image Translation, Machine Intelligence(DNN and Reinforcement Learning), Advanced Machine Intelligence, Cognitive algorithm and Computer Supported Interaction, Digital image processing

Volunteer work

The Earth Saviour Foundation*(India)*

ActionAid*(India)*

Redcross*(India)*

Other Interests

- Playing Guitar
- Playing Chess
- Traveling
- Cycling
- Swimming
- Reading Books

Oct 2007 - Jun 2011

Bachelor of Engineering Bengaluru, India
B. M. S. College Of Engineering
Computer Science
Object oriented modeling and design, Programming the Web, Java and J2EE, Unix System programming, Database management systems

PROJECTS

- **Trading algorithm using deep reinforcement learning**:- An application to trade electricity using Statistical analysis, Neural network(Encoder-Decode/CNN/RNN), and Reinforcement Learning(PPO/A2C).
- **Forecasting and volatility modeling of Timeseries data**:- A model to predict and model the volatility of gas consumption and weather data to be used for the Germany Energy market. Uses SARIMA, AutoArima, and Neural Prophet for forecasting and Arch/Garch models to find Values-at-risk.
- **Real-time Anomaly Detector on time-series data**:- A real-time anomaly detector application to detect inaccuracies in energy price data received from traders. Used Generative Model for novelty detection and got an accuracy of 97% with real-world data.
- **Real-time 3D Pose estimation using Synthetic Training Data**:- Uses the concept of 3D augmentation to create synthetic training data combined with a neural network to label and then estimate the pose of 3D objects Perspective-n-point algorithm in real-time.
- **News Article tagging : NLP**:- Developed a model to tag and classify news articles based on their title and content. Used One-hot-encoding and TFIDF techniques along with Deep Neural Networks.
- **Image translation : Pix2Pix**:- Models InfoGAN with cGAN uses the architecture of U-Net neural network for image translation which has been proven more accurate and faster than other baseline models.
- **Recommender system - Online Learning**:- A Collaborative filtering algorithm for the Recommender system using Python. The data contains the user and movie ratings acquired from Coursera.com. This project applies Online learning for training and generalizing.
- **Face recognition - Eigenfaces**:- Python code implementing Face Recognition based on Eigenfaces Algorithm. Data is from ATT and TU Berlin for result validation.
- **Multimodal Interaction**:- Unity3D/C project to control a Car in a game mode using multi-modal commands, speech, and keyboard/mouse using C and Unity3D.

</> Thesis

Implementation of an automated pipeline for random keypoint detection and evaluation for visual object localization on synthetic and real data
Fakultät Elektrotechnik und Informatik, TU Berlin
A Unity-3D(C#) and Python based pipeline to evaluate the pose estimation of 3D objects in real-world environment by using domain-randomized synthetic data to train Convolutions Neural Network and further using Perspective-n-point and Iterative closest point algorithms to estimate the pose. Results are bench-marked against datasets from YCB/Homebrew/TYOL.

ACHIEVEMENTS

Winnings:

Won two data challenges in a competition conducted by **DataHub Ruhr** in association with **Netconnect Germany**.

- Real-time anomaly detection of Gas consumption data
- Forecasting and volatility modeling for Weather and energy consumption data.

Certificates:

- Machine Learning (*Coursera, Feb 2019*)
- Deep Learning Specialization (*Coursera, Sep 2019*)