

Savin Shynu Varghese

Data Scientist & Machine Learning Engineer

Semmelweisstraße 11, 01159, Dresden | Born 27/04/1992 in India
+1 505 485 92 94 | savin.svarghese.2704@gmail.com | [LinkedIn](#) | [Github](#)
Check out my [Portfolio](#) here. I worked in the US for over 8 years and just relocated to Germany with my wife. I hold a valid Blue Card/work permit.



TECHNICAL SKILLS

Python (6 years of project experience) | R (2) | Machine Learning (4) | Pandas (6) | Numpy (6) | SciPy (6) | Matplotlib (6) | Big Data (6) | LLBrosa (1) | Jupyter (6) | ETL (3) | Deep Learning (2) | OpenCV (2) | Computer Vision (1) | Tensorflow (2) | Pytorch (1) | Keras (1) | Scikit-Learn/sklearn (4) | XGBoost (1) | Predictive Analytics (6) | NLTK (1) | Spacy (1) | Hugging Face (1) | Time Series (6) | Data Analytics (6) | Statistical Analysis (6) | Hypothesis Testing (5) | A/B Testing (2) | Anomaly Detection (6) | Digital Signal Processing (6) | Data Visualization (9) | PySpark (1) | Hadoop (1) | SQL (2) | MySQL (2) | NoSQL (1) | MongoDB (1) | AWS (2) | Docker (5) | Linux (10) | Git (5) | CI/CD (1) | FastAPI (2) | Streamlit (1) | NLP (1) | Spark MLlib (1) | Tableau (2) | Trello (2) | Jira (2) | Agile/Scrum (4)

WORK EXPERIENCE

06/2024 – today

Machine Learning Engineer

Omdena, California, USA (remote)

Project: AudioShield - An AI platform to detect deepfake audio.

- Developed deep learning models to detect deepfake audio by collaborating with a diverse team of data scientists, analysts and ML engineers.
- Built end-to-end deep learning (CNN and ResNet) models which can classify audio data with 0.93 accuracy and 0.93 F1-score to significantly reduce the deepfake audio problems experienced in Bangladesh.
- Deployed the CNN models for English and Bengali languages as web applications in the AWS EC2, Streamlit cloud and Hugging Face.

Technologies used: Python | Deep Learning | AWS | Docker | Streamlit | Huggingface | Signal Processing | Pandas | Numpy.

Project: Heart Disease Prediction - Web application to predict the possibility of coronary heart disease in patients.

- Developed an ML classification model (logistic regression, decision trees, random forest, XGboost, AdaBoost, support vectors and dense neural networks) with sci-kit-learn and Tensorflow to identify patients with a high risk of coronary heart disease.
- Deployed the fine-tuned model as a web application in Hugging Face.

Technologies used: Python | ML | DL | Exploratory Data Analysis | Pandas | Numpy | Matplotlib | Hugging Face | Scikit Learn | Tensorflow.

07/2021 – 05/2024

Research Data Scientist

SETI Institute, California, USA

Project: [COSMIC](#) - US research organization developing a computer cluster to search for extraterrestrial alien signals real time using radio telescopes.

- Collaborated with a diverse global team of scientists, software engineers and analysts to design and develop the real-time data processing computing cluster at the VLA telescope to search for intelligent alien life.
- Coordinated the development of an ETL pipeline to extract alien signal candidates from the 3TB/s of time series data in real time and stored them in an SQL database.

- Queried data from SQL database and utilized K-means clustering to classify 115,0000 alien signal candidates reducing manual inspection by 90 %. Conducted code reviews, and published the codes in GitHub and technical details as memos and scientific publications.

Technologies used: Python | Numpy | Matplotlib | Scikit Learn | Pandas | SQL | MySQL | Linux | Machine Learning | Clustering | Time Series Analysis | ETL Pipeline | Statistical Analysis | Anomaly Detection | Digital Signal Processing | Software Development.

08/2016 – 06/2021

Data Scientist

University of New Mexico, Albuquerque, USA

Project: **LWA Imager** - US-based organization building radio telescopes and data processing pipelines to monitor for astrophysical anomalies.

- Developed a Python-based image processing pipeline incorporating data cleaning, object detection, time series analysis, event classification, and anomaly detection to monitor energetic explosions.
- Analyzed 120 TB of LWA images, leading to the discovery of a new astrophysical explosion and 32 large meteors. Upgraded the pipeline with object tracking, boosting candidate classification accuracy to 95%, and discovering 86 additional large meteors.
- Applied principal component analysis to explore correlations in meteor features, enabling statistical analysis of their origins. Leveraged high-performance computing to parallelize the processing of 8 TB of OVRO-LWA images.
- Developed predictive regression models to forecast telescope response and source brightness for real-time calibration. Created Python tools to analyze 4 TB of time series data from the LWA telescope, using data cleaning, time series, and spectrogram analysis to detect short timescale astrophysical explosions.

Technologies used: Python | Numpy | Matplotlib | Scikit Learn | Pandas | Linux | Machine Learning | Clustering | Image Processing | Time Series Analysis | Anomaly Detection | Digital Signal Processing

EDUCATION

08/2016 – 12/2020

Doctor of Philosophy / Ph.D., Physics
University of New Mexico, Albuquerque, USA

08/2010 – 06/2015

Integrated Bachelors and Masters in Science / BS-MS, Physics
Indian Institute of Science Education and Research, Bhopal, India

CERTIFICATIONS

[Data Science Orientation](#) | [Tools for Data Science](#) | [Machine Learning Specialization](#)

LANGUAGE SKILLS

English (C2) | Malayalam (Mother tongue) | Hindi (C2) | German (A1, currently learning)

HOBBIES

Hiking | Football | Cricket | Watching Netflix

