VARUN GURUPURANDAR

+46 73754 0940 | vargu125@student.liu.se | Linköping, Sweden | LinkedIn

SUMMARY

Master's graduate in Data Science and Information Engineering from Linköping University, Sweden, with proven experience designing end-to-end ETL pipelines, engineering time-series data, and building production-grade machine learning models on complex industrial datasets. Proficient in Python, SQL, PyTorch, and scikit-learn across the full data lifecycle — from raw ingestion and cleaning to modelling, evaluation, and structured reporting. Experienced with containerised deployments and collaborative development. Published IEEE researcher driven by a passion for turning messy data into actionable insight.

PROFESSIONAL EXPERIENCE

Toyota Material Handling , Mjolby, Sweden --- Master Thesis (01/2026-07/2026)

- Built an end-to-end ETL pipeline in Python to ingest, clean, and synchronise multi-source CAN-bus sensor data (1 Hz–500 Hz, 41 channels); applied Binary Segmentation (BinSeg, L2 cost) to detect 7 operational motion phases — validated at silhouette score 0.675 and Rand Index > 0.9998.
- Computed phase-level KPIs (energy, power, distance, speed) across all load and battery state combinations; trained a PyTorch Multi-Task DNN with physics-informed loss achieving R^2 = 0.969 / 0.920 / 0.839 across acceleration, deceleration, and constant-speed phases.
- Developed a Hybrid CNN–LSTM for post-run energy estimation (R^2 = 0.986 / 0.983 / 0.912, 78 test trips) and a GMM safety classifier (BIC/AIC validated) to automatically flag unsafe driving segments.

EPIFAI , Data Engineering Internship (11/2025 – 03/2026)

- Engineered scalable ETL pipelines for ingesting, validating, transforming, and organizing unstructured enterprise data from multiple sources, improving data accessibility for downstream AI applications.
- Designed and implemented vector search infrastructure using LangChain, FAISS, and transformer-based embeddings, enabling high-performance semantic retrieval across large document repositories.
- Collaborated on the development of enterprise AI solutions leveraging Azure Search, OpenAI technologies, Docker, and modern software engineering practices to automate knowledge discovery, content standardization, and business reporting workflows.

EGNYTE , India -- Engineering Internship (08/2023-10/2023)

- Acquired good knowledge of machine learning frameworks, including PyTorch, TensorFlow, and Keras, with a strong focus on efficient API design and development.
- Gained hands-on experience with FastAPI for building high-performance APIs and Docker for containerization, enabling scalable and deployable machine learning solutions.
- Collaborated in a team setting with Git/GitHub for version control and CI/CD pipelines.

EDUCATION

Master's in Data Science and Information Engineering (08/2024 - Present)

Linköping University, Sweden

Relevent Coursework: Machine learning, Natural Language Processing, Statistics

Project: Analyzing and Mitigating Fairness Issues in NLP Models

- Analyzed bias in NLP toxicity classification models using fairness metrics on the Jigsaw dataset.
- Reduced model bias through data augmentation, improving fairness and prediction reliability

Bachelor's in Artificial Intelligence and Machine Learning (06/2020 – 07/2024)

Visvesvaraya Technological University (VTU), India

Project 1: IoT-Driven Temperature Data Analysis and Linear Regression for Enhanced Exploration (03/2022 - 08/2022)

- Implemented and configured IoT devices for real-time temperature data collection. Conducted exploratory data analysis (EDA) to identify patterns and trends using data processing tools such as Pandas, NumPy.
- IEEE International Conference on Data Science and Network Security (ICDSNS-2023), India, 2023,

Project 2: Exploring the Cosmos with Machine Learning: Advancements in Astronomical Data Processing.(11/2023 -04/2024)

- Developed an RCNN-based model for astronomical object classification using a dataset of 16,697 images of stars, galaxies, and celestial objects.
- Achieved 93.83% classification accuracy by leveraging deep learning techniques for automated celestial object identification.

2024 Second International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2024.

TECHNICAL SKILLS

Languages **: Python, SQL, Java**

Data Engineering **: ETL Pipeline , Data Ingestion & Validation ,Time-Series Analysis, Change-Point Detection**

Machine Learning & Analytics: **Regression, Classification, Clustering, CNN, LSTM, EDA, NLP**

Databases**: PostgreSQL** FAISS, Vector Search

Visualisation & Reporting: Power BI, Excel

DevOps **: Docker, Git, GitHub, FastAPI, CI/CD, Databricks**

Environments**: VS Code, PyCharm, Anaconda, RStudio, Linux, Windows**