Cannes, France
gagan.shan7@gmail.com
+33766356099
LinkedIn: linkedin.com/in/gaganjotshan/
GitHub: github.com/gaganjotshan
Portfolio: gaganjotshan.github.io/Portfolio/

Gaganjot Kaur SHAN

Junior Data Analyst

SUMMARY

Results-driven Data Scientist with 1.5 year track record in **machine learning**, **model development**, and **data analysis**. Proficient in **Python**, **SQL**, **Machine Learning frameworks**, **data visualization**, **MLOps**, and **system optimization**. Skilled in **feature engineering**, **algorithm selection**, and delivering **actionable insights** through **advanced analytics**.

EDUCATION

Masters of Science in Data Science and Engineering, *GPA: 3.62/4* — EURECOM, France

- *Coursework Highlights: Cloud Computing, Databases, Machine Learning, Deep Learning, Bayesian Statistics, Semantic Web, Image Processing, Mobile application and services, Security and privacy for big data and cloud*

Bachelors of Engineering in Computer Science, *CGPA: 7.95/10* — Chandigarh University, India

- *Coursework Highlights: Big Data Analytics, Data Warehouse and Data Mining, Information Retrieval, Genetic Programming, Design and Analysis of Algorithms*

TECHNICAL SKILLS

Programming: **Python**, **SQL**, **JavaScript**, **HTML**, **CSS**, **Shell Scripting**, **C++**, Core **Java**, **MATLAB**
ML/NLP: **Transformers**, **BERT**, **GPT**, **Word2Vec**, **spaCy**, **NLTK**, **Topic Modeling**, **TensorFlow**, **PyTorch**
Tools Used: **Docker**, **Kubernetes**, **MLflow**, **CI/CD pipelines**, **Power BI**, **Tableau**, **GCP**, **AWS (EC2, S3, Lambda, SageMaker)**

EXPERIENCE

Data Science Intern, iNeuron, Remote — Jul 2024 - Present

- **Engineered ETL pipelines** using **Python** and **PostgreSQL**, **improved financial record processing** by **5%**
- **Constructed interactive Power BI dashboards** to **visualize KPIs** impacting cost management
- **Building** a **machine learning based anomaly detection system** to monitor irregular patterns

Security Research Intern, SAP Labs, France — Mar 2023 - Aug 2023

- **Developed next-gen analytics application** employing **SAP Fiori (UI5)** to deliver secure **data analytics** across 5+ organizations
- **Diagnosed transaction discrepancies** and **rectified timestamp misalignments** in the **distributed ledger**
- **Deployed containerized solutions** with **Docker** on **Kubernetes** (Kyma clusters) to ensure **100%** availability
- **Configured YAML** files, **automated** and **debugged processes** with **Shell scripting** and employed **Postman** for **API testing**

Software Engineer Working Student, Indian Railways, India — Jun 2018 - Aug 2018

- **Created Python sockets application**-layer to **optimize data transmission latency**, enhancing system responsiveness by **3%**
- **Collaborated** with **cross-functional teams** to refine project strategies and improve outcomes
- **Presented insights** to stakeholders, influencing a **10-12%** boost in project alignment

RESEARCH PROJECTS

Real-Time AI Voice Tech Support, iNeuron — Aug 2024 - Present

- **Developed AI Voice Assistant** using **Large Language Models** (RAG and Mistral)
- **Improved response time** by **10-20%** by integrated **Qdrant vector database**

Industry-Standard MLOps: Practical Showcase — Apr 2024 - Jun 2024

- Executed **end-to-end ML pipeline**: data ingestion, transformation, model training, evaluation, deployment, **reduce rollout time by 25%**
- **Developed RESTful APIs** with **Flask** for ML **model integration**, **enhancing** project **modularity** and **scalability**
- **Orchestrated deployment** with GitHub Actions, **CI/CD pipelines** on **AWS**, **ensuring robust monitoring and version control**

Oscillatory Neural Network for Voice Spoofing Detection, EURECOM — Sep 2022 - Feb 2023

- **Implemented coRNN model** to effectively detect voice spoofing on ASV Spoof 2019 logical attacks database
- **Optimised model** architecture with **bi-directionality**, reducing **equal error rate to 6.8%** demonstrating **problem-solving**
- **Enhanced** detection **performance** by **1.9 percent point improvement**, **outperforming baseline** models.

Event Causality Detection using NLP, EURECOM — Jan 2022 - Jun 2022

- **Built Bidirectional GRU neural network** for event detection, **achieving 85% accuracy** on SemEval-2010 Task 8
- **Enhanced feature extraction** using **word2vec**, capturing **semantic relationships** from **10,000+ records** through **NLP techniques**
- Employed NLTK and spaCy, Gensim, and Keras and Scikit-learn for **data processing**, **word embedding**, and **model development**

Mathematics in Machine Learning, Indian Institute of Space Science and Technology — May 2017 - Jun 2017

- **Built** and **optimized machine learning models** for **predictive maintenance**, applying **logical and critical thinking**
- **Improved** equipment failure **forecasting** accuracy by **2-5%** through **feature engineering** and **ensemble methods**
- **Communicated impactful insights, analysed over 50,000 data points** utilizing **SQL and visualizing trends** in **Tableau**