Gaganjot Kaur SHAN

Cannes, France | gagan.shan7@gmail.com | +33766356099 | LinkedIn | GitHub | Portfolio

COMPUTER SCIENCE ENGINEER

Results-driven Data Scientist with 1.5 years experience in **machine learning**, **predictive modelling**, **data analysis**. Proficient in **Python**, **SQL**, **ML frameworks, visualization**, and **system optimization**. Skilled in **software engineering** and delivering **actionable insights**.

EDUCATION

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

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

Bachelors of Engineering Computer Science and Engineering, *CGPA: 7.95/10* — CU, India

- *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, Generative AI, LLMs**

Tools Used: **Docker, Kubernetes**, **MLflow**, **CI/CD pipelines**, **Power BI**, **Tableau, GCP, AWS (EC2, S3, Lambda, SageMaker)**

EXPERIENCE

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

- Rotation: Implementing ETL processes and financial data analysis to enhance cost efficiency and drive data-informed decisions.
- Processing and analyzing the data to identify key metrics and relationships, enabling effective cost management strategies.
- Tech & Skills: **PostgreSQL, Python, Git, Power BI, Web Scraping, Data Analysis, KPI Development, Documentation.**

Data Science Security Research Intern, SAP Labs, France — March 2023 - August 2023

- Developed a next-generation analytics application with integrated ledger technology and REST APIs, enhancing UX and enabling secure cross-organizational data synergy.
- Containerized with Docker and deployed on Kyma clusters, implementing auto-scaling, auto-healing for portability and reliability.
- Tech & Skills: **SAP Fiori (UI5), Docker, Kubernetes, JavaScript, Node.js, Flask, YAML, Postman, Shell Script, Debugging.**

Software Engineer Intern, Indian Railways, Patiala — June 2018 - August 2018

- Developed a custom application-layer class using Python sockets, optimizing data transmission latency.
- Achieved a 3% improvement in system responsiveness as a result of the enhancements.
- Tech & Skills: **Socket Programming, Performance Tuning, Software Engineering, Presentation, Cross-function teamwork.**

PROJECTS

Real-Time AI Voice Tech Support — August 2024 - Present

- Developing AI-Tech Assistant using RAG and Mistral, improved response time by 30-40% with Qdrant vector database integration.
- Implementing Generative AI and LLM techniques to enhance practical skills in advanced machine learning applications.

Industry-Standard MLOps: Practical Showcase — April 2024 - June 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 seamless ML model integration, enhancing project modularity and scalability.
- Orchestrated deployment with GitHub Actions and CI/CD pipelines on AWS, ensuring robust monitoring and version control.

Oscillatory Neural Network for Voice Spoofing Detection, EURECOM — September 2022 - February 2023

- Implemented coRNN model to effectively detect voice spoofing on ASV Spoof 2019 logical attacks database.
- Optimized model architecture by introducing bidirectionality, enhancing detection capabilities.
- Improved performance by 1.9 percentage points, achieving 6.8% equal error rate.

Event Causality Detection using NLP, EURECOM — January 2022 - June 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 through NLP techniques.
- Utilized Keras, NLTK, spaCy, Gensim, and Scikit-learn for data processing, word embedding, and model development.

PwC Switzerland Job Simulation - Power BI, Forage — June - July 2024

- Utilized analytical skills to develop Power BI dashboards, effectively conveyed KPIs, addressing client's data visualization needs.
- Demonstrated analytical problem-solving and strong communication by identifying and delivering actionable insights.

BCG Data Science Job Simulation, Forage — March - April 2024

- Conducted customer churn analysis using Pandas and NumPy, optimized model achieving 87.6% accuracy.
- Created data visualizations to interpret trends and prepared executive summary with actionable insights for Associate Director.

Mathematics for Machine Learning, Indian Institute of Space Science and Technology — May - June 2017

- Predictive analysis: Improved equipment failure forecasting accuracy by 2-5% through machine learning modeling.
- SQL, SVM, XGBoost, Random Forest, sklearn, Tableau, Predictive Modelling, Statistical Analysis, Advanced Mathematic.