

Naveen JOHNSON VALLAVANATT

Nice, France ▪ Email: naveenjohnsonv@gmail.com ▪ Phone: +33 6 02 56 18 12 ▪ LinkedIn: in/naveenjv ▪ Portfolio: naveenjohnsonv.fr

SUMMARY

Data Scientist (MSc, EURECOM) with 2+ years building analytics solutions for global enterprises across pharma, agrochemical, and energy sectors. Shipped LLM pipelines optimized for latency and reliability. Project expertise in Python, PyTorch, and Generative AI. I learn fast, document my work, and take autonomy, from prototype to production.

EXPERIENCE

Data Science Research Intern - AI Planning / NLP

ALTEN Labs

March 2025 - August 2025, Valbonne, Alpes-Maritimes, France

- Built an LLM pipeline (GPT-4o) converting natural-language narratives into Planning Domain Description Language (PDDL), executed with Fast Downward planner to produce playable interactive stories
- Added caching (state hashing) and automated schema checks with self-correction loops, cutting execution failures by 60%

Associate Consultant - Data Analytics

Blueprint Technologies Pvt Ltd

July 2021 - August 2023, Bengaluru, Karnataka, India

- Delivered 20+ dashboards with complex KPI logic for agrochemical, energy and pharmaceutical companies
- Led end-to-end ETL pipeline creation and backend data modelling; built checks across reporting layers, ensuring accuracy
- Oversaw user acceptance testing (UAT) and coordinated weekly client status reviews across 5+ enterprise projects

EDUCATION

Master of Science in Computer Science, Data Science Specialization

EURECOM ▪ Biot, France

2023 - 2025

- Focus: Machine Learning, Deep Learning, Reinforcement Learning, Advanced Statistical Inference, Distributed Systems

Diploma in Data Science

Indian Institute of Technology, Madras ▪ Chennai, Tamil Nadu, India

2021 - 2024

Bachelor of Technology in Electronics and Communication Engineering

National Institute of Technology, Calicut ▪ Kozhikode, Kerala, India

2016 - 2020

PROJECTS

Knowledge Engineering & Hallucination Reduction

EURECOM ▪ 2024

- Reproduced and extended the Text2KGBench framework, fixed codebase issues and integrated Odeuropa sensory dataset
- Benchmarked GPT-4o vs. Qwen2.5 32B on knowledge graph/ontology extraction from 16K+ sentences (DBpedia, Wikidata, Odeuropa datasets); GPT-4o achieved AUC 0.89, 99% conformance, 1% hallucination, outperforming all baselines

Sentiment Analysis with BERT

EURECOM ▪ 2024

- Fine-tuned transformer model BERT on 10K+ tweets for 3-class sentiment classification, applying text preprocessing and class-weight balancing; achieved 78% macro F1 score vs. lexicon-based VADER 64% and rule-based TextBlob 60%

Procurement Optimization and Profit Maximization for a Decorative Paint Retailer

IIT Madras ▪ 2022

- Analyzed ₹23M+ sales/₹9M inventory data; identified 80/20 revenue split, ₹900K dead stock, Dec/Mar seasonality
- Delivered actionable recommendations adopted by the business owner: demand-aligned procurement, CRM-driven loyalty plays, targeted pricing for slow movers, and supplier renegotiation using dead-stock data

SKILLS

ML / AI: Python, PyTorch, scikit-learn, HuggingFace, LangChain, GPT-4o, BERT, prompt engineering, RAG, Knowledge Graphs

Data: SQL, Pandas, PySpark, ETL pipelines, SAP Analytics Cloud, MongoDB, GCP (Vertex AI), AWS, Docker, Git

Languages: English - Proficient (C2), French - Intermediate (B1)