

Naveen JOHNSON VALLAVANATT

Nice, France • E-mail : naveenjohnsonv@gmail.com • Tél : +33 6 02 56 18 12 • LinkedIn : in/naveenjv • Portfolio : naveenjohnsonv.fr

PROFIL

Data Scientist (MSc, EURECOM) avec 2+ ans d'expérience en solutions analytiques pour entreprises mondiales (pharma, agrochimie, énergie). Pipelines LLM optimisés pour latence et fiabilité. Expertise en Python, PyTorch et IA générative. Apprentissage rapide, documentation rigoureuse et autonomie, du prototype à la production.

EXPÉRIENCE PROFESSIONNELLE

Stagiaire Recherche Data Science - Planification IA / NLP

ALTEN Labs

Mars 2025 - Août 2025, Valbonne, Alpes-Maritimes, France

- Développement d'un pipeline LLM (GPT-4o) convertissant des récits en langage naturel en langage de description de domaine de planification (PDDL), exécuté avec le planificateur Fast Downward pour produire des récits jouables
- L'ajout de la mise en cache et du hachage d'état avec boucles d'auto-correction a réduit les échecs d'exécution de 60 %

Consultant Associate - Data Analytics

Blueprint Technologies Pvt Ltd

Juillet 2021 - Août 2023, Bengaluru, Karnataka, Inde

- Livraison de 20+ tableaux de bord avec logique KPI complexe pour sociétés agrochimie, énergie et pharma
- Conception ETL et modélisation données de bout en bout ; contrôles sur les couches de reporting pour garantir l'exactitude
- Gestion des tests d'acceptation utilisateur (UAT) et coordination des revues hebdomadaires client sur 5+ projets entreprise

FORMATION

Diplôme Master en Informatique, mention Data Science

EURECOM • Biot, France

2023 - 2025

- Focus : Machine Learning, Deep Learning, Reinforcement Learning, Inférence Statistique Avancée, Systèmes Distribués

Diplôme en Data Science

Indian Institute of Technology, Madras • Chennai, Tamil Nadu, Inde

2021 - 2024

Diplôme d'Ingénieur (B.Tech) en Électronique et Télécommunications

National Institute of Technology, Calicut • Kozhikode, Kerala, Inde

2016 - 2020

PROJETS

Ingénierie des Connaissances et Réduction des Hallucinations

EURECOM • 2024

- Reproduction et extension du framework Text2KG Bench ; correction du code et intégration du dataset sensoriel Odeuropa
- Benchmark GPT-4o vs Qwen2.5 32B sur extraction graphes de connaissances/ontologies à partir de 16K+ phrases (DBpedia, Wikidata, Odeuropa) ; GPT-4o : AUC 0.89, 99% conformité, 1% hallucination, surpassant toutes les baselines

Analyse de Sentiments avec BERT

EURECOM • 2024

- Fine-tuning BERT sur 10K+ tweets pour classification sentiment 3 classes (prétraitement, équilibrage des classes) ; 78% macro F1 vs VADER 64% et TextBlob 60%

Optimisation des Achats et Maximisation des Profits pour un Revendeur de Peinture

IIT Madras • 2022

- Analyse de ₹23M+ ventes/₹9M stocks ; identifié répartition 80/20 des revenus, ₹900K stock mort, saisonnalité déc/mars
- Recommandations actionnables adoptées : approvisionnement aligné demande, CRM fidélité, tarification ciblée pour articles à rotation lente, renégociation fournisseurs sur données stock mort

COMPÉTENCES

ML / IA : Python, PyTorch, scikit-learn, LangChain, GPT-4o, BERT, prompt engineering, RAG, Graphes de connaissances

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

Langues : Anglais - Courant (C2), Français - Intermédiaire (B1)