

THIRUMURUGAN KUMAR

Villejuif, France +33 06 05 52 73 58

rishithiru1998@gmail.com | GitHub: thirurishi | LinkedIn: thirumurugan-kumar

Master's student in Data Science at EPITA seeking a 6-month internship as a Data Scientist. Strong expertise in machine learning, deep learning, and NLP, with experience in developing classification systems, optimizing models through cross-validation, and deploying ML APIs in containerized environments. Accustomed to transforming complex data into robust predictive models with measurable impact.

TECHNICAL SKILLS

- Programming:** Python (Pandas, NumPy), SQL
- Machine Learning:** Logistic Regression, Random Forest, Cross-Validation, Hyperparameter Optimization
- Deep Learning:** LSTM, GRU, TextCNN, Neural Networks,
- NLP & LLM:** TF-IDF, BERT, RAG, FAISS, LangChain
- Deployment & MLOps:** FastAPI, Docker, REST APIs, Containerization
- Databases:** PostgreSQL, SQLite, MongoDB
- Tools:** Git, Jupyter Notebook, Conda

EXPERIENCE

Android Application Development Intern

Dec 2022

National Small Industries Corporation (NSIC)

- Applied structured problem-solving and optimization techniques in real-world development tasks, strengthening analytical and debugging skills.

PROJECTS

Employee Performance and Attrition Prediction System

2025

- Developed a complete machine learning pipeline from 5,000+ HR records to predict employee attrition.
- Implemented and optimized (cross-validation) supervised models: Logistic Regression, Random Forest, and SVM.
- Final model performance: Accuracy 85-92%, Precision 88%, Recall 83%, F1-score 0.85.
- Deployed via FastAPI (<200ms per prediction) with batch processing of 10,000+ records/minute and automated monitoring (99.9% availability).

Fake News Detection Platform (Deep Learning & RAG)

2025

- Designed a full-stack platform (Streamlit, FastAPI, PostgreSQL) for automatic classification of news statements.
- Trained Deep Learning models (LSTM, GRU, TextCNN) on the LIAR dataset using an ensemble approach to improve prediction robustness.
- Implemented a complete NLP pipeline and integrated a RAG system using FAISS and LLM (LangChain) to generate contextual explanations.
- Integrated a BERT classifier to evaluate the relevance of evidence obtained through real-time web scraping; fully containerized using Docker.

EDUCATION

Master's in Computer Science – Data Science & Analytics

2024 – 2025

EPITA, France

Bachelor's in Information Technology

2020 – 2024

St. Joseph's College of Engineering, India

CERTIFICATIONS

- Python – IIT Bombay
- Advanced Google Analytics – Google
- BEC Vantage – Cambridge University Press

ACHIEVEMENTS

- Runner-up – Conocithon 2021 (IEEE)
- Finalist – Darkathon 2022

LANGUAGES

English (C1) | French (A2) | Tamil (Native)

CENTRES D'INTÉRÊT

Football | Photography | Chess