Navaraja Mannepalli
Data Engineer | DataScientist

Data Science & Analytics MSc Student | Paris, France | +33 745505726 | mannepallinavaraja@gmail.com

LinkedIn: linkedin.com/in/navaraja-mannepalli | GitHub: github.com/navaraja20 | Portfolio: https://personal-portfolio-zeta-sepia-50.vercel.app/

EDUCATION

Master of Science in Data Science & Analytics

EPITA School of Engineering and Computer Science, Paris, France (Sep 2024 – Jun 2026 expected)
Relevant Coursework: Machine Learning, Deep Learning, Big Data Analytics, Statistical Modeling, Data Mining, AI Fundamentals

BachelorofComputerApplications

KBN College, India (2020 – 2024)

PROFESSIONAL EXPERIENCE

Process Associate – Google Ads Optimization , HCL Technologies, Hyderabad, India (Jan 2023 – Jun 2023)

- Analyzed 50,000+ daily Google Ads campaign metrics using SQL and Excel to identify underperforming keywords and ad groups.
- Automated weekly performance reports using Python scripts, reducing manual reporting time by 70%.
- Collaborated with campaign managers to optimize bid strategies, improving ROI by 18% across 12 client accounts.
- Conducted A/B testing on ad copy and landing pages, resulting in 12% increase in click-through rates.

Amazon Intern , Online, India

- Developed customer segmentation model using K-means clustering on transaction data of 10,000+ users.
- Built Pipelines using Jenkins and Docker using Shell tools like Putty
- Presented findings to senior management, leading to targeted retention campaigns.

Machine Learning Intern, iNeuron Intelligence Pvt LTD, India

- Created fraud detection prototype using logistic regression and random forest on synthetic transaction data.
- Achieved 92% accuracy in identifying fraudulent patterns through feature engineering and cross-validation.
- Documented model performance and limitations in technical report for knowledge transfer.

PROJECTS

Self-Correcting RARE – https://github.com/navaraja20/Self_correcting_RARE (Final Project) - DL,LLM

- Developed and integrated a self-correcting biomedical question-answering system using retrieval-augmented generation (Llama-3.1-8B), SciBERT-based error detection, and T5-based answer correction pipelines, enabling high accuracy and reliability for medical queries.
- Implemented real-time error detection and correction workflows, including confidence scoring, correction validation, and audit logging in a HIPAA-compliant Streamlit web application.
- Collaborated on the end-to-end deployment of advanced analytics dashboards, batch processing features, and medical specialty filters, enhancing usability and compliance with healthcare data standards.

Employee Performance Prediction System – github.com/navaraja20/Employee-performance-prediction

- Built regression model to predict employee productivity scores using 6 months of HR data.
- Engineered 15+ features including attendance patterns, project completion rates, and feedback scores.
- Deployed model with 86% R-squared accuracy using scikit-learn and pandas.

Credit Card Defaulter Prediction – github.com/navaraja20/Credit_Card_Defaulter_main

- Developed classification pipeline to predict payment defaults with 89% AUC-ROC.
- Applied SMOTE for class imbalance and feature selection using mutual information.
- Visualized risk profiles using seaborn and matplotlib for stakeholder presentation.

Data Science Portfolio – github.com/navaraja20/dsp_navaraja_mannepalli

- Comprehensive collection of data cleaning, EDA, and ML workflows in Jupyter notebooks.
- Demonstrated reproducible research practices with Git version control.

TECHNICAL SKILLS

Subjects: Data Engineering, ETL/ELT, Big Data, ML Models, BI, Databases, Cloud, Web
Coding Languages: Python, SQL, Cypher Query, HTML, CSS, JavaScript, PHP
Tools: Azure, AWS, BigQuery, Databricks, PowerBI, Dataiku, Airflow, Docker, Grafana, Git, GitHub, Hugging Face
Libraries: NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch, TensorFlow, NLTK, Streamlit, FastAPI
Databases: MySQL, PostgreSQL, MongoDB, Neo4j, RedisData Visualization: Matplotlib, Seaborn, Tableau, Power BI, Excel