Arnav Kumar

+91 7038606744 | LinkedIn | Github | arnavu7038@gmail.com | Portfolio Website | Mumbai, India

PROFILE SUMMARY

Experienced Data Professional with a strong focus on deriving actionable insights and optimizing processes using Python/SQL. Proven ability to translate complex data into clear business solutions, automate workflows, and collaborate effectively with cross-functional teams.

EDUCATION

DIPLOMA IN DATA SCIENCE | CGPA : 8.2
IIT, MADRAS
2022

B.E, Electronics & Telecommunication | CGPA : 7.0
SMT, KASHIBAI NAVALE COLLEGE OF ENGINEERING, PUNE
2020

EXPERIENCE

Data Operations Analyst, Recro
11/2024 – 01/2025 | Bengaluru, India
- Executed ETL processes to integrate 3,000+ daily/monthly financial transactions into a SaaS portfolio system.
- Developed SQL scripts for efficient data extraction and manipulation.
- Automated data validation and error fixing using Python, improving report accuracy by 15%.
- Analyzed and resolved portfolio discrepancies during corporate actions with 100% accuracy.
- Streamlined workflows by combining ETL, SQL, and Python automation to meet contract goals.

Data Analyst, Cointab
05/2024 – 08/2024 | Mumbai, India
- Assisted in financial data reconciliation and process automation tasks for Cointab's client, Gameskraft, supporting their fraud detection workflows.
- Automated multi-stage financial reconciliation workflows using Python and SQL, improving accuracy and reducing manual effort by **100+ hours/month**.
- Configured the in-house tool using Python and SQL to automate reconciliation logic for financial audits, mapping transaction flows to identify points of friction.
- Coordinated with marketing and sales to create presentations, assess potential client needs, develop BRDs, Excel Models, and generate reports.

Data Science Intern, Happymonk AI Labs
05/2023 – 03/2024 | Remote, India
- Built 98% accurate object detection models (YOLO/Faster R-CNN) by aggregating multi-source data.
- Led large-scale data annotation (100000+ images) with rigorous QA checks to optimize model accuracy.
- Deployed real-time detection systems on 50+ concurrent video streams (recorded & YouTube Live) achieving <200ms latency with 95%+ inference accuracy.

Research Intern, Healthcare Technology Innovation Center
06/2023 – 09/2023 | Remote, India
- Contributed to ML/AI-based techniques for vascular health assessment using ARTSENS, an innovative image-free ultrasound system for arterial aging analysis.
- Developed Python scripts for automating ultrasound data screening and generating motion mode images, improving efficiency in data analysis workflows.
- Applied signal processing and machine learning skills to design and evaluate models for carotid artery wall dynamics, collaborating with a multidisciplinary biomedical research team.

TECHNICAL SKILLS

Languages & Tools
Python, SQL, Excel, Power BI, Git

Libraries & Frameworks
Pandas, NumPy, Matplotlib, Seaborn, re, scikit-learn, Scipy, Tensorflow, Pytorch

Machine Learning
Neural Networks, SVM, PCA, Linear/Logistic Regression, Random Forest, Naive Bayes

KEY SKILLS

ETL Pipelines • Predictive/Prescriptive Modelling • Generative AI • Data Visualization • Analytical Problem Solving • Stakeholder Communication

CAPSTONE PROJECTS

Document Intelligence Platform | Live Demo
03/2025 – 05/2025
- Developed SmartDoc AI, a full-stack platform integrating **RAG**, transformer AI, and advanced prompt engineering for document summarization and QnA.
- Engineered a secure FastAPI backend with JWT authentication for robust user, document, and API key management.
- Delivered intuitive user/admin dashboards with real-time analytics via Streamlit; containerized the app with Docker Compose for scalable deployment.

Boosting Gym Membership Engagement | Code Repo
07/2022 – 10/2022
- Analyzed 100K records using Python to identify high Retention clients and flexible schedule clients.
- Created interactive dashboards to visualize revenue drops, overcrowding trends,and ineffective decision making.
- Implemented promotions for data driven targeted strategies, directly improving client retention by **15%** and reducing overcrowding by **15%**.

OPEN SOURCE CONTRIBUTIONS

AI App for Forecasting Malaria | Live Demo
ML Engineer, Omdena
01/2024 – 04/2024 | Remote
- Collaborated internationally to develop an AI app for malaria forecasting, researching and acquiring data.
- Built and Optimized **XGBoost** models for malaria prediction, contributing to an expected **12-15%** reduction in deaths in **Liberia, West Africa**.
- Deployed the app via Streamlit and containerized with Docker for scalability.