ANJALI

+91 9716104186 anjaliduhanofficial@gmail.com Anjali Duhan anjali-anjalii

EDUCATION

Indian Institute of Technology Hyderabad (2022 - 2024)

- Master of Science in Physics: **CGPA - 8.4**
- Secured **AIR 405** in **JAM** conducted by IIT Roorkee in 2022.
- Extracurricular: Placement Coordinator, Clean India Drive Volunteer.

Gargi College, University of Delhi (2018 - 2021)

- Bachelor of Science in PCM: **CGPA - 8.7**
- **First Division** with distinction in Bachelor of Science.
- Extracurricular: Enactus Member, NSS Volunteer, Class Representative.

SKILLS

- **Programming Languages**: Python, SQL.
- **Machine Learning**: Numpy, Pandas, Sklearn, Matplotlib, Seaborn, PyTorch, Tensorflow, Streamlit, LLMs.
- **Data Science**: Data cleaning, visualization, modeling, hyperparameter tuning, and hypothesis testing.
- **Tech-stack**: LaTex, Github, Microsoft Office, Overleaf, VSCode, Pycharm, Jupyter, and Google Colab.

WORK EXPERIENCE

Data Scientist and Professor at Geeks of Gurukul (Sept 2024 - Present)

- Mentored projects — **from data cleaning to deployment** using real-world datasets and industry practices.
- Designed and led project-based learning modules, including **Movie Recommendation System, Financial Fraud Detection, and Sentiment Analysis**, to enhance the hands-on experience and project portfolios.
- Delivered **150+ training sessions** on Python, Data Science and Machine Learning, mentoring **200+ students**.
- Built Random Forest model with **78% accuracy** to predict student performance and support early interventions.

Data Science Intern at Stembridge Labs (May 2023-July 2023)

- Identified key purchasing patterns and trends in customer behavior by analyzing a year of transaction data, driving insights for targeted marketing and sales strategies.
- Developed XGBoost model to forecast purchase behavior with **85% accuracy**, improving sales conversion rates.
- Provided actionable dashboards and visual reports to business stakeholders using Python and PowerBI.

PROJECTS

PDF Chatbot with Conversational Memory (Github)

- Built an interactive chatbot capable of answering user queries from uploaded PDF documents in real time, enabling seamless document-based Q&A for end users.
- Implemented using OpenAI's GPT model for response generation and FAISS for fast vector similarity search.
- Developed using LangChain for workflow orchestration and Streamlit for the user interface, with OpenAIEmbeddings powering semantic vector search.

Predicting Fraud in Financial Payment Services (Github)

- Developed a fraud detection model to identify suspicious transactions in financial datasets with high accuracy.
- **Achieved an AUPRC of 0.99** by handling class imbalance & optimizing model performance for fraud cases.
- Used XGBoost, SMOTE, created 3D plots, and correlation matrices to highlight fraud patterns and key features.

Machine learning in High Energy Physics (Github)

- Built a machine learning pipeline to classify subatomic particles (Kaons, Pions, etc.) using Belle II detector data.
- Improved **particle identification accuracy** by **15%**, boosting the reliability of downstream physics analysis.
- Efficiently processed 1M+ data entries from six advanced detectors (SVD, CDC, TOP, ARICH, ECL, KLM) using ROOT and optimized XGBoost through extensive hyperparameter tuning.

RELEVANT COURSES

- **Academic Courses**: Calculus, Matrices, Algebra, Statistics, Probability, Differential Equations, Vector Calculus, Computational Techniques, Data Science Analysis, Machine Learning and its applications, and LaTex.
- **Non-Academic Courses**: Data Structures and Algorithms(Udemy), Programming with Python (Harvard's CS50), Introduction to ML (by Sebastian Raschka), Python Bootcamp, Python for DS & ML (Jose Portilla)