Yash Thakar

+91 8291006737 | yashthakar2710@gmail.com | linkedin.com/in/thakaryash | github.com/yashyaks

Education

B.Tech Computer Science & Engineering - Data Science [Honors in Computational Finance],
DJSCE, University of Mumbai (CGPA: 8.13/10)   Expected 2025
HSC, 12th Board Examination, Nirmala Memorial Foundation Junior College (96.5%)   2021
SSC, 10th Board Examination, Thakur Vidya Mandir High School (93.2%)   2019

Experience

Data Engineering Intern - Pay with Ring   Jun 2024 - Aug 2024
- Developed Python scripts to facilitate the migration of an existing data warehouse from Snowflake to ClickHouse.
- Enhanced a production-grade tool with AWS Lambda Functions, improving error handling and operational efficiency.
- Implemented a scheduled AWS Lambda function with Airflow to manage a proof-of-concept warehousing cluster, reducing off-hour operational costs.
- Created dashboards of AWS billing data, providing clear insights into high-cost areas for cost optimization.

AI/ML and Back-end Intern - Lawyantra   Nov 2023 - Feb 2024
- Introduced 60+ subcategories for automated case classification in Lawyantra's Gen-AI platform.
- Developed a Lawyer Ranking System architecture and integrated it with the Django REST framework and MySQL.
- Migrated the company's file storage system from AWS EC2 instance store to AWS S3, mitigating data loss risks.

Projects

AirViz - Big Data Pipeline
- Developed a data pipeline for ETL to transfer data from Amazon RDS (PostgreSQL engine) to an AWS Redshift data warehouse using AWS Glue and a spark runtime in AWS.
- Visualized airline flight schedules, airport connections, and airline delays and cancellations using PowerBI.

Olist's E-commerce insights
- Unveiled insights and trends from the Olist data and derived customer segments using RFM analysis.
- Developed ML models for customer segmentation, sales prediction, churn classification, and sentiment analysis.

Phishing Detector
- Developed a neural-network-based Phishing Link Detection model that uses 111 features, DNS data, and queries whois data for accurate link classification. Along with image similarity analysis to examine lookalike website structures.

PJM regional electricity consumption - Time Series modeling
- Researched the efficacy of time series models like ARIMA, SARIMA, GARCH, VAR, and CNN models for power consumption analysis. Uncovered trends and factors for energy policy and resource optimization.

Hybrid Quantum-Classical CNN for Audio Classification
- Developed a QC-CNN architecture using the Qiskit Machine Learning library to perform audio classification using an explainable mel-spectrogram approach on different dataset sizes and compared performance with classical models.

Skills

Languages: Python, SQL, C
ML and AI: Scikit-learn, TensorFlow, Pytorch, Langchain, Qiskit, Time Series Analysis, Deep Learning, CV, NLP
Data Science: Excel, Tableau, PowerBI, PySpark, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Selenium
Frameworks & Databases: Django, Flask, MySQL, PostgreSQL
Others: AWS, Git, GitHub, Postman, Netlify, Streamlit, Figma

Awards & Achievements

Retirement Industry AI Hackathon: T3 Hack, TIAA
Our team reached the semifinals and achieved a **Top-20 ranking nationally** at the Retirement Industry AI Hackathon: T3 Hack, a national-level competition **organised by TIAA** with over **2000 participants**.

Extra-curriculars

Chairperson - DJSCE Compute [djscompute.in]   Jun 2023 - Jun 2024
Lead a team of 20+ experienced mentors catering to a select batch of 60+ mentees to learn and build Machine Learning, Webdev, and App Dev projects.

Academic Mentor - We Can We Will Foundation [wecanwewillfdn.org]   Jun 2023 - Aug 2023