SHUBHAM RAUT

Pune, Maharashtra, India

+91-9405235277 | contactshubhamraut@gmail.com | linkedin.com/in/contactshubhamraut | github.com/realshubhamraut

SUMMARY

- Strong command over Python, Machine Learning, SQL, PowerBI and AWS enabling efficient data manipulation and analysis.
- Certified by Stanford, Google, and Harvard in Machine Learning, AWS, and Analytics with strong ML, statistical foundations.

EDUCATION

Lovely Professional University, LPU — Punjab, India

MCA, Masters of Computer Applications (Major in Data Science and Minor in Finance) — Aug 2022 – Dec 2024

RELEVANT COURSEWORK - Machine Learning, Data Structures, Deep Learning, Natural Language Processing

EXPERIENCE / INTERNSHIPS

Data Science Apprentice | Sparks Foundation Org. — Aug 2024 – Oct 2024

Analyzed recent Olympics 2024 data using Python, SQL, NumPy, Pandas, Machine Learning, and Git.

- Engineered and implemented data analysis **pipeline** by establishing connections with a **sqlite3** database, and developed a Python notebook using Pandas and NumPy by cleaning, analyzing and **relating 14** different datasets exceeding **10K** records.
- Conducted extensive exploratory data analysis (EDA) to identify trends, patterns, anomalies, and visualized findings with charts categorized by 3 medal types, leading to insights such as relationships between philosophical stance vs medal won.

PROJECTS

Query Cloud SQL in Natural Language | Python, Langchain, Langgraph, Snowflake, PostgreSQL, Streamlit, Gemini View Live

- Developed an intelligent **schema introspection engine** that automatically **constructs optimized SQL queries**, reducing manual query errors by **95%** and delivering interactive visualizations (e.g., any kind of charts) with real-time performance.
- Architected a robust data connectivity layer using SQLAlchemy and **LangChain** that seamlessly integrates cloud-based **Snowflake** and on-premises PostgreSQL with the help of Google Gemini API to dynamically generate and execute SQL queries , delivering sub-second query response times (under 500ms) and a unified, chat-driven LLM analytics experience.

Flight Airfare Prediction on Azure cloud with WebUI | FastAPI, Docker, Joblib, Scikit-Learn, Azure Cloud, Streamlit View Live

- Developed and deployed the containerized FastAPI microservice via Docker and Azure Container Registry, serving a flight airfare prediction model capable of handling **1000+** concurrent requests, thereby improving real-time access to predictions.
- Performed EDA and advanced feature engineering on flight data to streamline data preprocessing, built and benchmarked multiple ML models with integrated custom feature encoding and achieved over **90% prediction accuracy**—resulting in reusable encoder and feature transformation objects for efficient production through endpoints which exposed via FastAPI.

Advanced Medical diagnostic system with Flask on Azure | Flask, PyTorch, TorchVision, Docker, Github Actions View Live

- Engineered and deployed a highly scalable Flask-based web app on Azure using GitHub Actions with full CI/CD. Leveraged Docker to containerize deep learning models and the entire python environment, achieving an **end-to-end deployment pipeline** that processes industry-grade requests with over **99.9%** uptime.
- Integrated deep learning models with PyTorch and Scikit-Learn, validated on datasets exceeding 5,000 samples, and achieved diagnostic prediction accuracy of over 95%. integrated an ensemble of 3 deep learning architectures, reaching over 97% accuracy on key diagnostic tasks across 1,000+ test cases, then deployed them using docker on Azure Blob Storage creating possibility for instant further access for weights and deep learning models across the different apps if needed.

Automated First Screener and AI Job Matching NLP WebEngine | Python View Live

- Engineered and deployed a highly scalable Flask-based web app on Azure using GitHub Actions with full CI/CD. Leveraged Docker to containerize deep learning models and the entire python environment, achieving an end-to-end deployment pipeline that processes industry-grade requests with over 99.9% uptime.
- Integrated deep learning models with PyTorch and Scikit-Learn, validated on datasets exceeding 5,000 samples, and achieved diagnostic prediction accuracy of over 95%. integrated an ensemble of 3 deep learning architectures, reaching over 97% accuracy on key diagnostic tasks across 1,000+ test cases, then deployed them using docker on Azure Blob Storage

TECHNICAL SKILLS

Languages and Tools: Python, SQL, Microsoft Azure, ACR, Azure ML, AWS, lightsail, EC2, Streamlit, PowerBI, Figma, GIT, GitHub.

Libraries and Frameworks: Numpy, Pandas, Flask, Matplotlib, Langchain, PyTorch, Scrapy, Scikit-learn, Spacy, NLTK, ML Algos.

Data Science Skills: Machine Learning, Natural Language Processing, Statistical Analysis, Data Analysis, Data Processing

Soft Skills: Problem-Solving, Instructing, Presentation and Storytelling, Team Work, Collaboration and Communication.

CERTIFICATIONS / ACHIEVEMENTS

- Deep Learning Specialization - Stanford University
- Machine Learning Specialization - Stanford University
- Google Advanced Data Analytics - Google Inc
- AWS Fundamentals - AWS