Shivam Sapru

shivam.sapru@ucdconnect.ie (+353)-0894040368 DeansGrange, Co. Dublin linkedin.com/in/shivam-sapru

PROFESSIONAL EXPERIENCE

Data Scientist Intern, Mavenir 01/2024 – 05/2024
(Python, PySpark, AWS, MFLow, ZenML, AirFlow, Kedro, Kubeflow, Jira) Gurugram, India

- Optimized the **DeepMove** model for real-time mobility prediction and integrated it into scalable data pipelines.
- Trained and fine-tuned **LSTM** and **Transformer** models on large-scale time-series data for accurate mobility forecasting.
- Leveraged **AWS S3** for managing and storing training datasets and model artifacts across the pipeline.
- Improved MLOps workflows using **MLflow**, **Kedro**, **ZenML**, and **Kubeflow**, reducing model deployment and experimentation time by **20%**.
- Orchestrated end-to-end ML pipelines using **Apache Airflow**, enabling automated scheduling and reproducibility of workflows.

Machine Learning Intern, Mavenir 06/2023 – 08/2023
(Python, RCA, SQL, Jira) Bengaluru, India

- Uplifted the forecasting accuracy by 11% for multi-horizon time series data, by engineering 4 new features.
- Applied PyRCA using causal graphs and metric-based topologies and identified anomalous metrics using ε-diagnosis.
- Applied advanced feature engineering and causal modeling to enhance marketing analytics, including uplift modeling for campaign effectiveness and Bayesian inference for attribution analysis.
- Collaborated with cross-functional teams using **Jira** to align deliverables and track development progress.

EDUCATION

M.Sc. in Computer Science Negotiated Learning, University College Dublin 09/2024 – present
Key Modules: Cloud Computing, Data Science in Python, Big Data Programming, Machine Learning, Quantitative Methods for Engineers, Advances in Wireless Networking, Intro to RD & SQL Programming, Deep Learning Dublin, Ireland

B. Tech in Computer Science with honors in Data Science, Manipal University 08/2020 – 07/2024
Key Modules: Data Structures and Algorithms, OOPs, Artificial Intelligence & Fuzzy Logic, Computer System Architecture, Design and Analysis of Algorithms, Big Data Analytics, Operating Systems, Computer Networks, Database Management Systems, Data Science and Machine Learning, Data Mining and Warehousing, Natural Language Processing Jaipur, India

SKILLS

Programming Languages
Python, SQL, Java, Git

Data Analytics and Visualization
Pandas, NumPy, Matplotlib, Seaborn, Plotly

Predictive Modelling & AI
Scikit-Learn, TensorFlow, Pytorch, NLTK, XGBoost

Database & Cloud
AWS, MongoDB, Docker, RestFul API, Fast API, PostgreSQL, SQL, Spark, DataBricks, BigQuery

MLOps & Deployment
Airflow, MLFLow, KubeFlow, ZenML, Kedro, Prometheus

PROJECTS

Subtitle Translator (Microsoft) 05/2025 – present

- Led database and cloud infrastructure engineering for a **cloud-native subtitle translation platform**, built in collaboration with three Microsoft software engineers.
- Designed and implemented a scalable **PostgreSQL schema** to manage users, projects, and subtitle metadata; **migrated the database from local PostgreSQL to Azure Database for PostgreSQL** for improved scalability and cloud integration.
- Built **FastAPI** services for subtitle upload, translation via **Azure AI Translator**, and persistent file management using **Azure Blob Storage**.
- **Dockerized** the backend to standardize development environments and ensure smooth deployment to **Azure Web App**.
- Contributed to setting up a **CI/CD pipeline using GitHub Actions**, enabling automated testing and continuous delivery.

https://github.com/ShivamSapru/CS-GitLab

Scalable E-Commerce Backend System 12/2024 – 02/2025

- Developed a high-performance, microservices-based e-commerce backend using Java, Spring Boot, and Spring Cloud. Implemented user authentication (OAuth2), product management, order processing, and secure payments (Stripe API).
- Utilized RabbitMQ for event-driven architecture, Redis for caching, and Docker & Kubernetes for scalable deployment.
- Integrated Spring Cloud Gateway for API management and Prometheus for monitoring. Designed for high availability, fault tolerance, and real-time data processing, ensuring seamless user experience and operational efficiency.

ReneWind, Applied Machine Learning | University College Dublin 09/2024 – 10/2024

- Reduced generator maintenance costs at ReneWind using ML models (XGBoost, Random Forest, Bagging) with class imbalance handling using SMOTE and cost-based tuning.
- Productionized models via pipelines, optimizing for a user-defined cost metric based on true/false positives and false negatives.

https://github.com/ShivamSapru/ReneWind/tree/master

CERTIFICATES

- Google Cloud Certification: Machine Learning Engineer
- Oracle- Oracle Database Foundations
- Cisco PCAP: Programming Essentials in Python
- Google- Data Analytics Professional Certificate
- Cisco CCNAv7: Comprehensive Overview of Network Architecture and Protocols
- Udemy - MLOps Bootcamp: Mastering AI Operations for Success
- Cisco CCNAv7: Advance Concepts in Switching, Routing, and Wireless Essentials