PUJAN THAPA

https://www.linkedin.com/in/iampujan | iampujan.github.io | iampujan@outlook.com | https://github.com/iampujan

Skills

Programming: Python(Pandas, Pytorch, Numpy, sklearn, Plotly, Streamlit, FastAPI, Flask, Django, NLTK, Pyspark, Selenium, pytest, Kafka, Airflow, Scrapy, Beautifulsoup, requests, Scipy, matplotlib, opencv, Langchain, Langraph, HuggingFace, sklearn), LLMs model API Integration, etc
Databases: MySQL, Postgres, MongoDB, Redis, Oracle
Tools: Git, Docker, Jira, BitBucket, Github, Visual Studio Code, Pycharm, Anaconda, System Administration (Unix/Linux), Postman, Jmeter, CircleCI, GitHub Actions, MLFlow
Cloud Services: AWS (S3, Cloudwatch, Lambda, Kinesis Data Streams, EC2, SQS, IAM, Sagemaker)
GCP(Cloud Firestore, BigQuery, App Engine, Cloud Functions, Cloud IAM), Azure (Container Apps, Container Registry, etc.)

Experience

ML Engineer | Omdena | *Remote* | **May 2024 - July 2024**

Identifying Potential Areas for Urban Agriculture in Milan, Italy: Researched and developed solution during the timeframe of 8 weeks for Omdena Challenge from the data collection phase using Google Earth Engine, exploratory data analysis, model development using both supervised and unsupervised learning and model deployment in Streamlit Community Cloud.

Software Engineer | Leapfrog Technology | *Kathmandu, Nepal* | **Dec 2020 - Aug 2023**

- Developed, deployed, and maintained projects utilizing Python, AWS Services (S3, Lambda, Kinesis, EC2, SQS, CloudWatch), Oracle, Docker, Git, and CircleCI with minimal supervision.
- Developed and monitored cron services using Python and Airflow.
- Re-architected XML file parsing and data storage for scheduled cron jobs using AWS Kinesis Data Streams Producer/Consumer architecture, AWS Lambda, and CloudWatch, replacing traditional methods.
- Performed unit testing, integration testing, and load testing for web platforms and REST APIs using pytest, Selenium, JMeter, and Automation Anywhere.
- Significantly enhanced test coverage from less than 30% to over 90%, ensuring comprehensive testing of the service and substantially improving its reliability and quality.

Software Engineer | Parentiv Inc. | *Remote* | **Dec 2019 - Dec 2020**

- Researched and developed the Parentiv app services architecture using FastAPI, Google Cloud Firestore, Bitbucket for version control, and deployed on Google App Engine.
- Implemented text preprocessing techniques for API responses using Python NLTK and Spacy.
- Designed and developed an API layer to connect the Cloud Firestore database with external data sources like Apple Watch and Google Home Assistant.
- Collaborated with the app development team to perform unit testing and regression testing of APIs and web backend using pytest and JMeter.

Education

- University of Naples Federico II, **Master of Data Science** — *Naples, Italy* — Sept 2023 - Sept 2025
- Tribhuvan University, **BSc. in Computer Science and Information Technology** — *Kathmandu, Nepal* — 2015-2019
- Fusemachines, **AI Fellow 2024**: Graduated as an AI Fellow for 6 months comprehensive AI Course and training for upskilling, providing hands-on experience and real world AI applications in areas such as machine learning, computer vision, NLP, Generative AI, LLMs, etc.
- Deep Learning Nanodegree — Udacity — 2019

Master Thesis

- **Failure Robust HD Maps for Autonomous Driving:** Developing a **novel multimodal pipeline** for High-Definition (HD) map construction that can flexibly incorporate various types of sensor inputs—including monocular/multi-view cameras, LiDAR, and radar. The research leverages **images and point clouds** to efficiently represent and reconstruct complex 3D environments. This approach aims to replace traditional handcrafted and sensor-specific pipelines with a **generalizable, data-driven framework** for scalable and consistent HD map generation across diverse driving scenarios.

Projects

- **Content Based Information Retrieval for Earth Observation):** Developed a Python-based system to retrieve satellite images with similar content to a query image by leveraging spectral and spatial feature analysis. Using PyTorch and a pre-trained ResNet-50 model, deep feature embeddings were extracted to represent image content, while NetworkX was employed to model image relationships and perform similarity searches within the database. The approach combined deep learning with traditional descriptors such as color histograms and texture patterns, enabling the detection of regions with comparable land cover, vegetation, and urban structures in large-scale Earth observation datasets(landcover.ai and SEN12MS).
- **CREA3 (In Progress):** CREA3 stands for Conflict Resolution with Equitative Algorithms funded by the European Union (EU) in

collaboration with University of Naples Federico II and other companies and universities.The CREA3 Project aims to modernize civil law procedures throughout the European Union, enabling citizens, legal professionals, judicial authorities, and, in particular, vulnerable users to achieve swifter, fairer, and non-discriminatory access to justice. Its strategic vision aims to augment the performances of its predecessors, CRE and CREA2, seamlessly introducing electronic signatures, authenticated videoconferences with automated evidentiary logging, and third, providing inclusive access measures that support individuals with limited digital literacy and cognitive capabilities.

- **Plant Disease Detection:** Developed a plant disease detection system using image analysis. Leveraged YOLO for object detection, alongside techniques like Non-negative Matrix Factorization (NMF) and Fuzzy Clustering for feature extraction and classification using Tensorflow, Roboflow, scipy, matplotlib, sklearn, ultralytics, etc.
- **RAG Chat Bot:** Developed the frontend and backend for European Commission News retreival using React JS and FastAPI. Integrated the GPT and GROQ API and embedding models with Langchain and quadrant as a vectorstore.
- **FATER - Developing Neighborhood Analytics Model capable of estimating the potential of point of sale in the category Baby Diapers in order to create geolocalized commercial plans:** Analyzed diaper market potential for Fater (P&G subsidiary) and refined revenue forecasts for Naples stores using socio-demographic, geographic, and points of interest data. Identified key factors like store type, parking availability, and temporal shopping patterns to optimize strategic marketing and revenue growth. Applied advanced statistical tests, geo-spatial analysis, and machine learning models with hyperparameter tuning to predict store performance.
- **Political Leaning Detection in News:** Developed a web application to classify news articles as left-leaning, centrist, or right-leaning using fine-tuned BERT models. Enhanced model efficiency through quantization and Low-Rank Adaptation (LoRA). Deployed the solution using React JS, FastAPI and Docker, hosted on Azure Container Apps for real-time predictions, supporting media bias detection and analysis.
- **US Stock Analysis:** Analyzed 157 US Energy stocks and performed advance clustering for identifying bullish and bearish trends revealing overbought and overpriced stocks, profitable investments, and stocks facing losses. Additionally, a focus on volatility analysis pinpointed the most volatile stocks in the dataset. This comprehensive approach enables investors to make informed decisions and manage risks effectively. Furthermore, seamless integration with Kafka ensures real-time updates and subscriptions for stakeholders, enhancing accessibility to critical market insights.

Awards and Extracurriculars

- **Co-organizer, Python Users Group Nepal Meetup**

Organized and volunteered for Python meetups and university workshops.

- **Volunteer, EuroPython 2025**

Volunteered as Session Manager for speakers

- **Volunteer, Pycon DE and PyData 2025**

Volunteered as Session Chair for speakers

- **Volunteer, Pycon Italia 2024**

Volunteered as Talk Manager and Runner.

- **Volunteer, EuroPython 2024**

Monitored Discord channels.

- **Microsoft Student Partner**

Planned and organized tech workshops (HoloLens, Global Azure Bootcamp, etc.).

- **Second Runner Up, Asian Hack 2019**

Developed “Airify”, a mobile application that gives air quality information about different cities and recommends solutions to stay safe from polluted air based on air quality index.