Rohail Taimour

Python Software Engineer | Statistician | AI/Machine Learning Specialist

1040, Brussels, Belgium
rohail.taimour@gmail.com | +32 489 83 64 76 | Personal website | Linkedin | Github

Education

MSc in Statistics - KU Leuven, Leuven, Belgium - 2014-2016

Graduated Cum Laude, Master’s thesis on continuous optimization of production processes in MATLAB

BSc (Hons.) in Accounting and Finance - Lahore University of Management Sciences (LUMS), Lahore, Pakistan - 2010-2014

Graduated with Distinction (3.6/4.0)
Courses: Operations Research, Supply Chain, Decision analysis, Applied Probability
Treasurer for University Adventure Society organizing hiking trips for groups of upto 300 people in North of Pakistan

Technical competencies

Programming Languages:	Python and R with 7+ years of experience
Cloud Services:	AWS (S3, EC2, ECS, Cloudwatch, SageMaker Studio), Azure (Blob, Databricks, Pipelines)
CI/CD:	Git, GitHub Actions, Azure Pipelines, GitHub CLI, Docker, Docker compose
Databases:	PostgreSQL, Redis, SQLite3, Neo4j, SQLAlchemy, Alembic
Monitoring:	Grafana, Sentry, Kibana
Code quality tools:	pytest, coverage, responses, factoryboy,

Summary

Seasoned Python Software Engineer with a Master's degree in Statistics. I am proficient in building applications following best software practices such as CI/CD, containerization and operating in an AWS or Azure cloud environment.

Job History

- **Python Backend Engineer** at Engie GEMS, Brussels, Belgium | *Jan 2024 - Present*
- **Python Software Engineer** at Illumina, Mechelen, Belgium | *April 2023 - October 2023*
- **Machine Learning Engineer** at GSK Vaccines, Brussels, Belgium | *October 2022 - Feb 2023*
- **Lead Data Scientist, AI Developer** at UCB BioPharma, Brussels, Belgium | *August 2016 - October 2022*

Relevant Experience

Backend Engineering for Green Energy Trading Platform using Flask and Python Ecosystem

Python Backend Engineer, Engie Green energy Management solutions (GEMS), Brussels, Belgium Jan 2024 - present

- Supported European green energy trading processes by being first line support for incidents on supported applications.
- Diagnosed and resolved issues with Okta token management, troubleshooting client credentials flow for integrating with an external system.
- Enhanced resilience of a scheduled **celery** task making an API call by incorporating retry mechanism. Extended our internal mock for testing celery task execution to support testing that tasks are appropriately retried by monkeypatching celery internals.
- Led the migration of two codebases from **SQLAlchemy** v1 to v2. Updated the contributing guidelines to include preferred querying syntax by conferring with the development team.
- Ensure full coverage of challenging features involving celery tasks triggered via SQLAlchemy after commit flush events, where test assertions changed depending on system time.
- Resolved a bug where a PR failing the test running step could still be merged into the main branch by adding additional checks within the pipeline and reducing code duplication by using reusable workflows when running with matrix strategy.
- Implement new endpoints for system integration between different services and diagnose issues using a mixture of **Kibana**, **Sentry** and issue reproduction using review environments.
- Designed and implemented via github actions an end to end integration system involving a database, a message queue, multiple external services that could be run in a scheduled manner.

Python solution enabling launching, pipeline management and automated report downloads for Customer-Uploaded Data

Python Software Engineer and Data Pipeline Architect, Illumina, Mechelen, Belgium April 2023 - October 2023

- Designed a Python service that automates the monitoring and processing of customer-uploaded sequencing data, initiating further analysis or report generation based on predefined criteria.
- Implemented a dual-layered approach: the first layer handles the initiation and tracking of analysis pipelines, while the second layer is registered as a **Docker** image in the analytics backend to perform post-processing on the output files and create comprehensive summary reports for the customer.
- Scheduled the Python service to operate every 30 minutes for new data and updates, ensuring seamless progression from data upload to final report delivery to customer environment.
- Implemented comprehensive systems integration, utilizing a combination of CLI tools and API calls for effective coordination and automation across various software components.
- Applied Object-Oriented Programming (OOP) principles to organize API, database interactions and endpoint processing to reduce code duplication and utilize self-documenting object names.

faker, pre-commit, ruff

Personal details

- Nationality: Belgian, Pakistani
- Languages: English (fluent/bilingual), Urdu (Native), French (B1)
- Mobility: Driving License available, flexible for hybrid setup in Belgium
- Availability: TBD
- Hobbies: Drumming and percussion instruments, Bouldering/Climbing, productivity, Squash, reading

- Implemented unit testing using **pytest** and implemented fail-safe mechanisms for robust error handling.

Automated SQL Script Generation to facilitate PostgreSQL Data Migration in multiple environments

Python Software Engineer, Illumina, Mechelen, Belgium July-Aug 2023

- Designed and implemented a data ingestion framework to parse and validate input files for generating and validating SQL update statements.
- Conducted comprehensive testing of generated SQL scripts using a mock of production database tables to test that SQL scripts run as expected.
- Utilized SQLAlchemy for database schema management, creating and populating mock tables in a test environment to ensure the integrity and functionality of SQL scripts.
- Implemented the solution as a Python package encapsulating the entire data migration logic within a Docker entrypoint for portability and ease of deployment.
- Leveraged **Jinja2** templating to generate dynamic, parameterized SQL scripts, enabling the script to adapt seamlessly across different deployment environments, such as Development, Integration, and Production.

Design and implement information retrieval methods using Natural language processing (NLP)

Machine Learning Engineer, IT Supply Quality, GSK Belgium Oct 2022-Feb 2023

- Improved performance of information retrieval by 20% on unseen test data using a custom named entity recognition (NER) from **Spacy**.
- Performed POC's on Azure DataBricks environment to improve model performance using rule-based techniques as well as **NER** and annotated data to train custom NER.
- Added text preprocessing features to the NLP pipeline such as **Spacy** tokenization, Part of speech (POS) tagging, better handling of non-english emails, breaking emails into sentences, etc.

Unit Commitment Solver for Power Grid Optimization via FastAPI

- Developed a REST API using **FastAPI** for optimizing energy distribution among powerplants based on load requirements and fuel costs.
- Implemented multiple algorithms to solve the **unit-commitment problem**, considering factors like fuel cost, powerplant efficiency, and environmental constraints.
- Utilized **Pydantic** for data validation and schema definition, ensuring data integrity and streamlined request handling.
- Packaged and containerized the application using **Docker**, with detailed documentation and a Dockerfile for easy deployment and scalability.
- Employed **pytest**, along with Python best practices such as typing and linting.
- Managed project dependencies using **Poetry**, facilitating efficient workflow and package management.
- Deployed the API service using **Uvicorn** and integrated a **Swagger UI** for interactive API documentation and testing

Yield optimization for batch and continuous production processes using Machine Learning in Python

Lead Data Scientist, Supply and Manufacturing, UCB Switzerland/Belgium Aug 2020-Oct 2022

- Production setting proposed by model directly led to an increased throughput of 20%, turning in a recurring 1.5 million euro in annual cost savings
- Analyze time series data collected from equipment sensors and visually summarize golden batch insights
- Created (Bayesian) and tree-based regression models to quantify impact of process changes and predict batch performance
- Performed a thorough model validation and hyperparameter tuning exercise before recommending model insights be tested in a live production environment
- Supported delivery of workshops demystifying the process of conducting AI projects and machine learning to process experts

Marketing Mix Optimization and Customer Segmentation modelling in

EU5

AI/ML engineer, Lead Data Scientist, Go to Market/Commercial EU5, US and Japan, UCB June 2019-June 2021

- Utilized demographic data to develop customer segmentation models, identifying key segments for high-potential growth and revenue.
- Integrated multi-channel marketing data, gaining a comprehensive understanding of customer interactions to inform the development of an effective market mix model.
- Leveraged the market mix model to investigate the relationship between customer segments and marketing channel responsiveness, pinpointing segments with the highest engagement potential.
- Aligned the insights from the market mix model with promotional strategies, targeting specific customer segments through their most responsive channels.
- Optimized marketing resource allocation based on the model's insights, tailoring marketing strategies to enhance customer engagement and maximize ROI across diverse channels.
- Adapted data science methodologies to address country and product specificities, delivering tailored solutions for up to ten different use cases across various products and countries.
- Developed a Python package with **Cookiecutter** templates that abstract the complexities of the data science workflow, enabling configurable deployments across diverse scenarios such as different countries and disease areas.
- Enhanced the package to seamlessly wrap over **scikit-learn**, thereby simplifying key data science tasks from preprocessing to model training and tuning
- Incorporated **MLflow** into the package for robust artifact management, allowing for the tracking of model versions, data inputs, and predictions
- Performed feature engineering using **PySpark** and validated ingested data using data visualization methods and discussions with subject-matter experts

Personal projects

Web Scraper to analyse Property Purchase and Rental Trends in Belgium

- Developed web scraper using Beautiful Soup to collect information such as apartment data such as price, area, etc.
- Implemented SQLite for data storage, using `**Pydantic**` for data validation and `**SQLAlchemy**` for database interactions.
- Encapsulated the concerns into a python package with dependency management using Poetry.
- Employed **Prefect** for job orchestration, managing the workflow's scheduling and monitoring of scraping tasks.

Personal Portfolio and blogging website built using Hugo and hosted using Github Pages

- Created website using **Hugo** and implemented features such as a contact form, and visitor commenting capabilities.
- Hosted the static website on GitHub Pages and automated the deployment process using GitHub Actions.
- Codebase hosted on github

Automated Resume Builder and Continuous Deployment System with GitHub Pages Hosting

- Engineered an automated system for generating, versioning, and hosting a dynamic CV using Markdown, HTML, Jinja templating and CSS.
- Set up a trio of GitHub repositories to separately manage the CV's content, styling, and public hosting on Github Pages.
- Developed a Python package for automating the styling and generation of the CV, integrating with Markdown and HTML/CSS.
- Implemented version control for CV content using a private GitHub repository, ensuring secure and organized data management.
- Leveraged GitHub Actions for automating the CV's generation and deployment process, enabling updates through git pushes.
- Hosted the final CV on GitHub Pages, providing a live, online version that can be easily updated