

Rohail Taimour

Contract Data Scientist | Statistician | Python/R for AI/Machine Learning
1040, Brussels, Belgium



rohail.taimour@gmail.com | +32 489 83 64 76 | [Linkedin](#) | [Github](#)

Summary

Experienced data scientist with an expertise in working on data products that have a machine learning focus. I am able to play a dynamic role between data scientist/machine learning engineer and data engineer given the needs of the project. I am passionate about adopting development best practices wherever possible and am comfortable working in an ambiguous and fast changing environment, laser focused on delivering user requirements in an agile manner.

Education

MSc in Statistics

KU Leuven, Leuven, Belgium

2014-2016

- Graduated Cum Laude, Master's thesis on continuous optimization of production processes in **MATLAB**
- Treasurer for University Adventure Society organizing hiking trips for groups of upto 300 people in North of Pakistan

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

Personal details

- **Nationality:** Belgian, Pakistani
- **Languages:** English (fluent/bilingual), Urdu (Native), French (B1)
- **Mobility:** Driving Licence available, flexible for hybrid setup with up to 4 days on site
- **Availability:** Immediately
- **Hobbies:** Drumming and percussion instruments, Boulderling/Climbing, productivity, Squash, reading

Freelance projects (Oct 2022-present)

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

Machine Learning Engineer, IT Supply Quality, GSK Belgium

Oct 2022-Feb 2023

- **Achievements:**
 - Refactored and simplified python codebase, adding features such as better handling of non-english emails, breaking emails into sentences, etc
 - Improved performance of information retrieval by 20% on unseen test data
- **Responsibilities:**
 - Implemented a text preprocessing pipeline using `spacy` tokenization, Part of speech (POS) tagging
 - Performed POC's on how to improve NLP model performance using rule based techniques as well as named entity recognition (NER)
 - Annotated data to train a custom NER
 - Defined and clarified KPI's to track model performance evolution

Data science projects at IT AI team, UCB Pharmaceutical (2016-Oct 2022)

IT Artificial Intelligence (ITAI) is a global and cross functional team working across different use cases in the pharmaceutical industry from commercial (go to market), finance, manufacturing, to drug development and adverse event reporting. Our projects used a modern stack consisting of Docker as a consistent development environment, Amazon AWS for remote servers, Microsoft Azure for CI/CD pipelines and git versioning.

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

Lead Data Scientist, Supply and Manufacturing, UCB Switzerland/Belgium

Aug 2020-Oct 2022

- **Achievements:**

- Supported delivery of workshops demystifying the process of conducting AI projects and machine learning to process experts
- 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
- **Responsibilities:**
 - Deliver a two-hour presentation weekly on insights from data, focusing on driving engagement and discussion within the process experts
 - Deliver biweekly sprint review update on deliverables status
 - Analyze time series data collected from equipment sensors and visually summarize golden batch insights
 - Supporting the mentoring and training of junior data science profiles on the project
 - Performed a thorough model validation and hyperparameter tuning exercise before recommending model insights be tested in a live production environment
 - Created (bayesian) and tree-based regression models to quantify impact of process changes and predict batch performance

Optimize customer engagement based on promotional responsiveness by channel across EU5

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

June 2019-June 2021

- **Achievements:**
 - Delivered as many as ten different use cases as lead data scientist for different products and countries
 - Created Python packages wrapping over scikit-learn facilitating the training and tuning of models as well as data pre-processing steps
 - Supported reproducibility of analyses using MLflow that could allowed junior profiles to deliver higher quality, reproducible analyses
 - Templated entire data science workflow into a configurable pipeline that could be executed for different countries/disease areas
- **Responsibilities:**
 - Facilitate timeline and feasibility assessment for data science component together with the project manager
 - Populate backlog with data science tasks during dailies and help prioritize to support timely deliveries
 - Investigated adaptations to data science methodology for country/product specificities for maximum reusability
 - Deliver biweekly sprint review update on deliverables status
 - Validated ingested data using data visualization methods and discussions with subject matter experts
 - Supported data engineers in the creation of features using pyspark
 - Created customer segmentation models and proposed optimal resource allocation based on customer responsiveness to different channels
 - Developed a framework for logging model versions, data inputs, predictions and other model artifacts to a dashboard to facilitate discussion with the team and our stakeholders

Scientific influencer (KOLs) identification, ranking and profiling using network analytics and Neo4j

Data scientist/Product owner, Drug Development, Commercial, Medical affairs, UCB

2018-2019

- **Achievements:**
 - Successfully delivered 10+ KOL identification projects to stakeholders across the company
 - Reduced time to deliver reports from days to hours by leading functional interface improvement with a senior software developer
- **Responsibilities:**
 - Introduce stakeholders to our product offering and understand their needs
 - Prioritized request backlog, make request feasibility assessments for customisations and communicate timelines
 - Define product roadmap to increase our capacity for taking in requests by improving the software foundations of our data product
 - Create custom databases for each new request, ingesting additional data sources to quantify influence, as needed
 - Performed data quality checks by querying the database using Cypher from the Web UI as well as jupyter notebooks
 - Made network visualizations using networkx, Cytoscape and performed custom analysis
 - Provided customized KOL ranking/profiling reports and presentations based on stakeholder requirements

Developed an automated forecasting workflow of claims data from US healthcare system

Lead Data Scientist, US Finance and claims, UCB

2017-2018

- **Achievements:**
 - Created modular R packages to extend the functionality of Facebook's prophet package
 - Created an end to end workflow for ingesting, forecasting and reporting with a user friendly interface to analyze forecasting results
 - Achieved forecasting accuracy of > 90% across the different use cases
- **Responsibilities:**
 - Prototyped different time series forecasting methods to flexibly model multiple time series models
 - Applied anomaly detection methods to account for outlying behavior in time series automatically
 - Performed hyperparameter tuning and validation on a batch compute machine
 - Created customized reports in Rmarkdown to show forecast metrics and visualization for each individual time series that was emailed to our stakeholders

Hands-on workshop to demystify Artificial intelligence and Machine Learning

- **Achievements:**
 - Delivered the workshop to over 100 people in four different venues and received great feedback on level of engagement
 - Created a R `shiny` application to facilitate conducting the workshop, walking people through the typical AI use cases
- **Responsibilities:**
 - Prepare content for the workshop and decide on most engaging way to deliver the material
 - Conduct the workshops, via hands on exercises on the content and debrief participants on solutions to the exercises

Technology stack and competencies

Development environment

- Preferred python packages: pandas, pymc3, kedro, scikit-learn, sktime, seaborn
- OS: Windows, MacOS, Linux (Redhat/Ubuntu)
- IDE: Pycharm, VScode, Rstudio, jupyter notebooks, Azure data bricks, Docker containers
- code quality: black, pre-commit, isort, typing
- git flow branching for collaboration using git
- AWS s3/Azure blob for project related data exchange
- Azure pipelines for CI/CD and testing
- conda, mamba, pip for project dependencies
- MLflow for experiment tracking, model and artifact management, serving predictions

Machine Learning models

- Regression, classification, clustering using classical statistical methods
- Machine learning approaches (tree-based methods such as XGboost, Random Forest etc) on tabular data
- Time series forecasting, clustering, aggregation using bary centers
- Probabilistic programming, Monte Carlo simulation and Bayesian statistics
- Natural language processing (NLP) text embeddings, Named entity Recognition (NER), sentiment/topic analysis

Soft skills

- Driven, proactive, dynamic, can-do attitude
- Diverse educational/work experience: accounting and finance, supply chain, statistics, data science
- Quick study for different business use cases, technology stacks, (machine learning) methods
- Strong communication skills and ability to deliver to deadlines
- Can work autonomously, as well as within a team
- Experienced in Agile rituals and breaking down complex requirements into achievable milestones

Certifications

- Python: 'Software engineering for Data scientists' (2019) from Ian Ozsvald (PyData London)
- R: 'Advanced R Programming' (2014) from Belgian Open Source Analytical Consultants (BNOSAC)

Personal projects

- Developed webscraper in python (Beautiful soup, Selenium) to compare apartments based on price, area, etc
- Created a predictive model for whether movie will achieve an imdb rating > 7.5 using best practices for data science projects