

Profile

James Twose is an experienced Data/Machine Learning Engineer with a proven track record of 7 years, specialised in Python based analyses. He is detail-oriented and driven and excels at utilizing data-driven insights to develop innovative solutions for complex challenges. He has experience in all aspects of the process, from initial ideation/stakeholder management to requirements scoping, infrastructure implementation, data engineering, data analysis, machine learning implementation, MLops and visualization/explanation of output to stakeholders.

His passion for data engineering and machine learning, and his dedication to staying at the forefront of industry trends, allows him to deliver efficient and accurate data/ML pipelines. Because of this breadth of experience, James is able to fit in at any point of the process, but he is happiest when he can be part of multiple facets of a project.

Contact

- **Amsterdam, The Netherlands**
- **English (native), Dutch, French**
- **Email:** contact@jamestwose.com
- **Phone:** +31 628465035
- **LinkedIn:** [James Twose - LinkedIn](#)
- **GitHub:** [James Twose - GitHub](#)
- **Website:** [James Twose - Website](#)



Technical Skills

- **AWS:** S3, EC2, Lambda, SageMaker, RDS, DynamoDB, Boto3
- **Cloud Services:** Microsoft Azure, Google Cloud Platform, Heroku, DataBricks, Digital Ocean, Netlify, Deta, Render, Clever Cloud, Vercel
- **Databases:** SQL, NoSQL, PostgreSQL, MySQL, MongoDB, SQLite
- **Infrastructure:** Docker, Docker Compose, Kubernetes, Terraform
- **Version Control:** Git, GitHub Actions, TektonCD, ArgoCD
- **Data Visualization:** Matplotlib, Seaborn, Plotly, Dash, Streamlit
- **Data Science/ ML:** Python, Pandas, Polars, Numpy, Scipy, SciKit-Learn, StatsModels, Pingouin, Pymer4, Semopy, Shap, CasADi, PyTorch, TensorFlow
- **Web Development:** HTML, CSS, JavaScript, ReactJS, VueJS, Django, Flask, FastAPI, Streamlit, Dash
- **R:** Tidyverse, Caret, Catboost, XGboost, Keras, Shiny, ggplot2, plotly
- **Julia:** DataFrames, VegaLite, MLJ, Flux, Statistics, PlotlyJS
- **Other Languages:** Bash, ZSH, LaTeX, Rust(personal use), Golang(personal use)
- **Other:** Visual Studio Code, Pycharm, SPSS, Qualtrics, Nvivo, Microsoft Suite, Jupyter Notebook, Jupyter Lab, RStudio, Spyder, Vim, Atlassian suite (Jira, Confluence, Trello)

Work Experience

Data Engineer - Event Connectors (Feb 2023 - May 2023)

At Event Connectors, I was hired to create pipelines and frontend code to build an internal dashboard, which is used to monitor one of Event Connectors' products, Feed Factory. Feed Factory is the interface used by municipalities throughout the Netherlands to organize and manage the activities and events in their respective areas. With the work I did, I was able to provide insight into the number of scheduled events managed through Feed Factory, and the associated amount of time saved by using this product.

Used CI/CD via github actions. Using Docker containers, this was how we delivered the APIs.

More specifically, my responsibilities included: - Setting up PostgreSQL backend (PostgreSQL/SQLAlchemy), retrieving data from various APIs (previous infrastructure used MongoDB and Elasticsearch)

- Creating middleware to communicate with the backend so that the frontend can easily and quickly retrieve data for figures and tables (FastAPI, Docker)
- Create cron jobs to automatically update the backend when new input occurs every day (SQLAlchemy, GitHub actions)
- Create plots and MVPs for frontend engineers to integrate into the main application (plotly.js, react.js)

Data Scientist - Citizen Shield (June 2021 - May 2023)

Citizen Shield is an initiative of the Finnish government to provide technological, behavioral and societal solutions for protective actions to tackle pandemics in Finland. My role within this project was to analyze and describe the results of a large nationwide survey aimed at understanding the behavioral determinants leading to the use of masks in Finnish culture. On this basis, research policy has been drawn up and/or adapted to better respond to pandemics in Finland in the future. For this project I acted as a Data Scientist, with the following responsibilities:

- Data Analysis (Python, Catboost, shap, statsmodels, pingouin, streamlit)
- Data Visualization (Python, streamlit, plotly, HTML, CSS)

- Writing an academic article

Data Engineer & Data Scientist - Neurocast (Apr 2021 - Apr 2023)

Supervising junior data scientists. Tasked with researching and creating tailor-made analytical methods to monitor and predict changes in Multiple Sclerosis (MS) based on data collected from smartphones. Inspiration for these approaches comes from complex systems science (e.g. Recurrence Quantification Analysis), optimal control theory (e.g. Licitra et al., 2018), electrical engineering (e.g. Single Input Single Output systems) and neuroscience (e.g. Subtype and Stage Inference).

In addition, we have created AI models in AWS Sagemaker for the following purposes. Anomaly detection: find out who the user of a phone is, yourself or someone else, activity classification: find out whether the user is walking, sitting or standing based on typing behavior. In addition, predicting symptom worsening (in this case in people with muscular disease MS) based on sensor data.

Used CI/CD via github actions.

Responsibilities: - Data Engineering - python (AWS S3, AWS RDS, boto3, pandas, polars, MySQL, pytest, GitHub actions) - Data Analysis - python, R (sklearn, PyTorch, pyRQA, CasADi, seaborn, matplotlib, plotly, dash, mlVAR, uSEM, pySuStain, Pingouin, Semopy, Catboost, XGboost) - Stakeholder management (jira, confluence, microsoft suite) - Supervising junior data scientists (jira) - Writing an academic article (microsoft suite) - Results presentation (HTML, CSS, JS)

Medior Data Engineer & Data Scientist - Neurocast (Apr 2020 - Apr 2021)

At Neurocast I was the lead data scientist appointed to perform analyses in collaboration with the Amsterdam University Medical Center. This collaboration, intended to validate the use of smartphone interactions as a means to monitor and predict outcomes in Multiple Sclerosis (MS), led to a cross-sectional journal article, a responsiveness journal article, and a longitudinal journal article.

Responsibilities: - Data Engineering - python (AWS S3, AWS RDS, boto3, pandas, MySQL, tsfresh, pytest, GitHub actions) - Data Analysis - python (sklearn, jupyter, plotly, pyme4,) - Stakeholder management (jira, confluence, microsoft suite) - Writing an academic article (microsoft suite) - Results presentation (HTML, CSS, JS, R, shinyapps, docker)

Data/ Machine Learning Engineer - Neurocast (Apr 2019 - Apr 2020)

Created a machine learning workflow and performed an assortment of machine learning analytics (SVM, Random Forests, Multilayer Perceptron etc.) to predict MS outcome measures using keystroke data. These analyzes were all internal, but an example of a similar approach is available upon request.

Responsibilities: - Data Engineering - python (AWS S3, AWS RDS, boto3, pandas, MySQL, tsfresh, SQLAlchemy) - Data Analysis - python (sklearn, MERF, jupyter, seaborn, matplotlib) - Stakeholder management (jira, confluence, microsoft suite) - Writing an academic article (LaTeX, Microsoft suite)

Junior Data Scientist - Neurocast (Apr 2018 - Apr 2019)

Applied non-linear time series analyzes - specifically dynamic complexity of keystroke dynamics to match fluctuations in keystroke data with clinically relevant changes in outcome measures. This resulted in the following journal article: 'Twose et al, 2020'. Also looked at using Mixed Effects Random Forests to predict MS outcomes, taking into account the hierarchical nature of clinical data and still applying a non-linear approach (as opposed to Linear Mixed models).

Responsibilities: - Data Engineering - python (AWS S3, AWS RDS, boto3, pandas, MySQL, SQLAlchemy) - Data Analysis - python, R (sklearn, caret, MERF, pyunicorn, seaborn, matplotlib, jupyter, tidyr, casnet) - Stakeholder management (jira, confluence, microsoft suite) - Writing an academic article (LaTeX, Microsoft suite)

Intern - College ter Beoordeling van Geneesmiddelen (Jan 2015 - Jan 2016)

Established a functioning relational database for medicines; examined vector autoregressions to predict adverse events; examined scientific advice to change the protocol that the MEB uses to regulate antidepressants.

Responsibilities: - Setting up a database (MySQL) - Data analysis (R, tsDyn, tidyverse, Nvivo) - Results presentation (microsoft suite)

Research Assistant - University College Roosevelt (Jan 2015 - Jan 2016)

Measuring perceived stress levels, coping strategies, and physiological correlates in UCR students during times of high and low workload: Selected to conduct this study for University College Roosevelt management to influence policy decisions about the protocol the university uses to manage stress at combat students.

Responsibilities: - Research design - Data Collection (google forms, qualtrics) - Data Analysis (SPSS) - Results presentation (microsoft suite)

Education

Radboud University (2016 - 2018)

Master of Science - Behavioral Sciences - Thesis - Moral ambiguity: how empathy and psychopathy inform moral decision-making

University College Roosevelt (2013 - 2016)

Bachelor of Science - Liberal Arts and Sciences - Major: Cognitive Science and Life Science - Minor: Psychology

Havant College, UK (2010 - 2012)

International Baccalaureate - Mathematics, French, English, Physics, Chemistry, Psychology