Othmane Bahraoui

www.linkedin.com/in/othmane-bahraoui/ | othmane.bahraoui@alumni.duke.edu

Results-driven data professional with experience in analyzing large datasets, building predictive models, and integrating machine learning pipelines. Skilled in statistical analysis, feature engineering, and data cleaning to uncover insights and drive decision-making. Expertise in working with machine learning techniques such as decision trees, feature importance analysis, and advanced predictive models. Proficient in conducting exploratory data analysis (EDA) and presenting findings through visualizations. Data Analyst and ML Engineer.

Education

Duke University — Durham, NC, USA

Bachelor of Science, **Data Science** — 2019 - 2023

Relevant Coursework: Advanced Data Visualizations, Algorithms and Databases, Advanced Linear Algebra, Probability and Statistics, Programming and Data Structures, Statistical Machine Learning, Numerical Analysis and Optimization.

Skills

Python · Java · R · SQL · MATLAB · PostgreSQL · MySQL · AWS · Pytorch · HTML · CSS · JavaScript · Git · GitHub · PowerBI · Tableau · UNIX Commands · MS Excel · MS Word · PowerPoint

Professional Experience

Software Engineer, AI Training Data (Freelance) — May 2024 - Present

Outlier AI — Remote

- Enhanced AI Model Performance and Accuracy: Developed and debugged complex Python and SQL prompts to train large language models, resulting in improved code generation accuracy and a significant reduction in errors.
- Designed and evaluated sophisticated prompts, optimizing the AI model training process and reducing the task completion time, while ensuring high-quality data annotation and analysis.

Data Scientist Intern — November 2023 - March 2024

NayaOne — London, UK

- Analyzed over 100 diverse financial datasets, driving insights for innovative fintech products.
- Generated synthetic datasets valued at over £40,000, leading to successful client sales.
- Led exploratory data analysis EDA initiatives, identifying over 50 key patterns, shaping strategic decisions.
- Developed ML predictive models for financial forecasting, risk assessment, and customer behavior analysis, improving accuracy by 25%.
- Conducted API testing for data-driven solutions, resulting in a 30% reduction in bugs and errors.

Languages

- English, French, Arabic (Native)
- Chinese (Intermediate)
- Spanish (Beginner)

Interests

- Photography
- Calisthenics
- Football
- Filmmaking

Certifications

Currently pursuing:

- Amazon Web Services AWS Solutions Architect Associate

Projects

Flask Based Image Classification Web Application

- Developed and deployed a Flask-based API on Render to classify images using a pre-trained CNN model with a 53.6% accuracy on the CIFAR10 dataset, and a responsive frontend built using HTML, CSS, and JavaScript.
- Used PyTorch for image preprocessing and model prediction, ensuring efficient cross-platform functionality and live demonstration capabilities. **Python, PyTorch, HTML, CSS, JavaScript, Flask**

Machine Learning API in AWS Sagemaker for Penguin Species Classification

- Built and deployed an XGBoost model using Amazon SageMaker to classify penguin species based on features such as bill length, flipper length, and body mass. Managed the full data pipeline, from preprocessing and feature engineering in Jupyter Notebook to data storage in S3.
- Deployed the model with AWS Lambda and API Gateway, enabling real-time predictions via a REST API. Utilized Python for API development, data exploration, and model training, ensuring a serverless architecture for scalable and efficient inference. **Python, AWS Sagemaker, lambda, S3, EC2**

Gamifying Risk Identification for Alcohol Use Behaviors

Bass Connections Independent Study

- Team Leader for the Data Analysis and Implementation sub team to code and model two techniques to identify alcohol use risk behavior: Balloon Analogue Risk Task and Brief Intervention.
- Developed code in both Python and MATLAB to implement BART (Balloon Analogue Risk Task): Balloon Analogue Risk Task, effectively analyzing and simulating risk behavior scenarios. **Python, MATLAB**

Weather Forecast Predictions

- Analyzed a weather dataset for various regions in Australia, employing Random Forest, Logistic Regression, and MLP neural network models to accurately predict temperature and next day rainfall.
- Evaluated model performance using standardized measurements, such as Mean Squared Error, Mean Absolute Error accuracy, and R2, ensuring robust and reliable weather forecast predictions and generating 84% accuracy. **Python**

Premier League Defense Analytics

- Developed an interactive Power BI dashboard to analyze Premier League defensive performance, with filters for season and team data, focusing on top 6 teams. Visualized key defensive metrics including wins, losses, saves, goals, dispossessed balls, clearances, tackles, and interceptions.
- Created detailed graphs and pie charts for clean sheets, penalties conceded and saved, allowing users to explore defensive patterns and trends across multiple seasons with flexible filtering options. **PowerBI**

Insurance Pricing Analysis

- Conducted comprehensive analysis of insurance pricing using four CSV datasets, consolidated data, handled duplicates, and missing values. Explored correlations between factors such as insured amounts and number of bedrooms/bathrooms through visualizations.
- Applied decision trees to assess feature importance, revealing pricing trends influenced by market dynamics and insured amounts. Analyzed partner pricing variations, providing strategic recommendations to enhance competitiveness. **SQL, Python**

Hybrid Movie Recommendation System

- Implemented a hybrid movie recommendation system combining content-based filtering, item-based collaborative filtering, and user-based collaborative filtering to suggest personalized movies. Calculated a weighted average score for each movie using user ratings and vote counts.
- Used TF-IDF and cosine similarity for content-based filtering based on genres and descriptions, KMeans clustering for item-based filtering, and user-based collaborative filtering to predict movies based on similar user preferences, creating a well-rounded recommendation system. **Python**