Emeline Floc'h

+1 845-770-7495 | efloch03@gmail.com | Linkedin | Github | Personal Portfolio

Experience

Senior Data Scientist August 2021 – Present

Gro Intelligence (AI Solutions for Agriculture and Climate) *New York, NY*

- Designed a Machine Learning framework to develop yield models for various crops, feature engineering three decades of daily satellite climate data, and identifying the best model architecture (Linear Regression, Gradient Boosting, Neural Network, etc.). Cutting down the yield model development cycle from multiple weeks to 2 days, beating industry-standard accuracy from Agricultural Departments.
- Built CatBoost times series predictive models to guide an international NGO's efforts in addressing food insecurity. Forecasting production volume of major grains in African countries, outperforming US Department of Agriculture forecasts with a 12% reduction in Mean Absolute Percentage Error.
- Established and maintained data quality protocols and alerts for our predictive models, ensuring ongoing accuracy and reliability throughout the agriculture seasons.
- Implemented advanced linear algebra techniques from recent research to stress-test global trade networks in response to major climate disasters, enhancing the method by integrating key new features.
- Led end-to-end ML solution implementation in Python with Docker, AWS EC2, AWS ECR, and Airflow, ensuring smooth code deployment and scheduling in production, demonstrating expertise in full data science project lifecycle. Spearheaded R&D projects within a high-growth environment, taking charge of project timelines and output specifications.

Data Scientist April 2018 – July 2021

Locus Analytics (Economics Research Firm) *New York, NY*

- Developed a robust data pipeline, employing Convex Optimization, to fetch, feature engineer, and impute over 120GB of Census time series data, serving as a foundational support for various Data Science projects
- Implemented an exponential smoothing time series model for employment and wage forecasting, surpassing Bureau of Labor Statistics ten-year projections by 6 million workers with a remarkable -5% Mean Absolute Percentage Error (MAPE).
- Created multi-class, multi-label Neural Network (Multilayer Perceptron) and Log-Linear classifiers to assign standardized network graph attributes to companies and jobs, extracting insights from unstructured data like public company annual reports and job postings.
- Applied comprehensive text processing techniques, including HTML parsing, feature engineering, and NLP methodologies such as Word2Vec and Topic Modeling, to handle unstructured data effectively.

Data Science Research Intern Jul. 2017 – Aug. 2017

SNCF (French National Railway Operating Company) - Research & Innovation *Paris, France*

- Applied time series forecasting models, including ARIMA and Gradient Boosted Trees, to predict train station foot-traffic, reducing Root Mean Square Error (RMSE) by 52% from baseline.
- Introduced Dynamic Time Warping to discern ridership patterns, cluster stations effectively, and improve the accuracy of forecasts.

Personal Projects

Script generation using Transformers | *Python, Pytorch*

- Implementation of Transformers from scratch to generate new TV show scripts trained on favorite shows

Citi Bike Share Prediction using Autoformer | *Python, Pytorch, HuggingFace*

- Applying Autoformer to forecast the volume of ride per day per bike sharing station

Education

Carnegie Mellon University Pittsburgh, PA

Machine Learning Masters Exchange Program *Sept. 2017 – Jan 2018*

Grenoble Institute of Technology - Ensimag Grenoble, France

Masters and Bachelor Degree in Computer Science and Applied Mathematics *Sept. 2013 – Sept 2018*

Technical Skills

Programming: expert in Python and SQL, familiar with Java and R
Data Science: pandas, numpy, seaborn, plotly, scipy, statsmodels, sklearn, PyTorch, HuggingFace, Tableau
Deployment: Git, Docker, AWS (EC2, ECR), Airflow, Django