Emeline Floc'h

+1 845-770-7495 | efloch03@gmail.com | linkedin.com/in/emeline

Experience

Senior Data Scientist August 2021 – Present

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

- Designed an automated tree-based Machine Learning framework to predict major crop yields, feature engineering three decades of daily satellite climate data.
- Built times series predictive models to guide an international NGO's efforts in addressing food insecurity. Forecasting production volume of major grains in African countries, achieving 6% deviation from historical ground truth, surpassing US Department of Agriculture forecasts by 12%
- 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 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.

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*

Personal Projects

Personal Research Project | *Python, Flask, ElasticSearch* January 2021 – Present

- Augmented database of Financial Services events to identify trends in the industry
- Built event classifier from text description with word embedding, Latent Dirichlet Allocation and SVM model
- Fetched data from ElasticSearch and built a Flask API to retrieve analysis results to back-end engineers

Technical Skills

Languages: expert in Python and SQL, familiar with Java and R
Data Science Libraries and Tools: pandas, numpy, seaborn, plotly, scipy, statsmodels, sklearn, TensorFlow, Tableau
Other Tools: Git, Docker, AWS (EC2, ECR), Airflow, ElasticSearch, Flask, Django