George Lewis

ghl227@gmail.com
(203) 918-4859

Summary: Data Scientist and Data Engineer with a passion for building data products, accurate and effective statistical models, data pipelines, and tools to facilitate valid and repeatable data analyses

Skills

Languages: Python, Java, SQL, Scala, Clojure. Limited experience with R. Past experience with C and C++

Tools: Scikit-Learn, Pandas, Scipy, Jupyter, Postgres, Redshift, BigQuery, Spark, Airflow, AWS, Weka, XGBoost, matplotlib, seaborn, Flask, Play, Ring. Limited experience with Tensorflow, Theano, tidyverse

Experience

LendUp *2013 - 2017*
Head of Risk and Analytics
Principal Data Scientist

- Founded the Risk and Analytics team, which later became the Data Science team. Designed and implemented the risk and underwriting program, growing the company's loan portfolio by a factor of X while decreasing loss rates over YY basis points. Led the team from seed round funding through series B.
- Designed, trained, deployed, and monitored statistical models using logistic regression, generalized additive models, decision trees, and random forests to evaluate the credit risk of all applicants and to identify and mitigate fraud.
- Grew the team from 1 to XX, establishing best practices for statistical analysis, techniques and tools for monitoring and real-time analytics. Enforced high software standards and helped to develop a culture of strong programming practices within the data science team
- Built production-quality software in Java to implement our risk program, to pull and parse data into features in real time, to score our statistical models, and to take action as a result of the risk decision, sending the customer to the desired next step in our flow
- Recruited and trained the original data engineering team. Worked with data engineering to build ETL processes for analytics, frameworks for running jobs using Airflow and Spark, pipelines for feature extraction, and tooling to enable streamlined model deployment for data scientists

Graduate Research *2008 - 2013*
New York University and The European Organization for Nuclear Research (CERN)
Member of the ATLAS experiment at the Large Hadron Collider (LHC)

- Developed statistical modeling and analysis software used extensively throughout a 3000 person experiment
- Used massively parallel batch computing to analyze petabytes of data distributed over a worldwide grid of data centers
- Designed and Implemented novel statistical modeling techniques to perform state-of-the-art inference
- Enforced coding standards, conventions, and best practices across a large C++ and Python code base
- Recipient of National Science Foundation US LHC Graduate Student Support Award

Education

PhD in Experimental High Energy Particle Physics *2013*
New York University, New York, NY

B.A. in Physics and Mathematics *2007*
Columbia University, New York, NY