Daniel Kleeman

github.com/dankleeman | linkedin.com/in/daniel-kleeman-7259a648
dan.kleeman@pm.me | https://dankleeman.github.io

Experience

Shipt - Partner Experience Machine Learning April 2022 - Present
Staff Machine Learning Engineer

- Own an end-to-end NLP project to classify product brands to grow $16 Million in CPG partner revenue sharing.
- Develop fine-tuned BERT-based multi-class model to identify candidate brands from 10 Million product names.
- Develop naive Bayes model to classify among 5,000 candidate brands with high precision and explainability.
- Own an end-to-end sparse product substitutions recommendation project to save $5 Million in order GMV.
- Deploy Bayesian model to meet product need for 45,000 daily product substitution events while researching GCNs.
- Contribute to team python library for standardized use of pre-trained HuggingFace/PyTorch models.
- Deploy containerized production models on airflow for batched training and prediction parallelism.
- Own improvements to CI/CD pipeline and enforcement of code quality in code review process.
- Mentor junior data scientists and help them grow with machine learning and software development knowledge.
- Participated in internal hackathon contributing hand-writing OCR to cart prototype; won CTO award.

Ørsted - US Trading Analytics January 2021 - April 2022
Senior Data Scientist

- Owned ERCOT-wide electricity load forecast model methodology and real-time production key trading input.
- Developed and deployed web app dashboards presenting real-time data analysis results to the trading team.
- Refactored analytics backend into a Python package for code reuse and more robust deployment.
- Implemented CI/CD pipeline to build, test, and deploy web apps within production and testing environments.
- Mentored a Junior Analyst on learning how to write code and build production-quality software.
- Acted as technical lead enforcing best practices including code reviews, unit testing, and a git branching strategy.

University of Chicago March 2020 – May 2023
Teaching Assistant

- Held office hours and graded for Applied Data Analysis, covering probability and statistical learning with emphasis on: linear models, clustering, neural networks, SVMs and tree-based methods using both boosting and bagging.

Morningstar
Quantitative Research - Quantitative Software Engineer October 2017 – January 2021

Morningstar Quantitative Rating for Funds (MQR) Cloud Re-platform

- Refactored legacy libraries to leverage parallelism in the cloud-based serverless architecture for 10x speedup.
- Refactored and replatformed ETL processes from SQL to Python utilizing Pandas, Numpy, and multiprocessing.
- Provided feedback and extensive code reviews to junior analysts and engineers.

Morningstar Quantitative Rating for Funds (MQR) 2.0

- Researched and developed an ML model to assign ratings to 40,000 mutual funds, settling on random forest.
- Co-authored a white paper evaluating efficacy of the model via event studies, as a primary subject matter expert.
- Replaced R- and SQL-heavy codebase with Python (mainly pandas) codebase, focusing on maintainability.
- Prepared an extensive suite of unit tests to ensure correctness of production code.

Buyside Sales Team - Associate Client Service Manager November 2015 – October 2017
Private Data Team - Data Analyst July 2014 – October 2015

Education

University of Chicago Graduation Date: March 2021
Master of Science in Computer Science

University of Illinois at Urbana-Champaign Graduation Date: May 2014
Bachelor of Science in Finance

Technical Skills

Software: Python, Pandas/SciPy/Scikit-learn, PyTorch, HuggingFace, Jupyter, SQL, Git, Linux
Cloud Tools: Athena/Presto, Snowflake, Airflow, Kafka, Kubernetes, Docker, AWS