Zachary Galante

(774) 277-1757 Medway, MA LinkedIn Medium Google Scholar ZGalante@Berkeley.edu

Education

University of California, Berkeley — Expected 2024
Master of Information and Data Science

Bryant University, Smithfield, RI — 2022
Bachelor of Science in Data Science

Skills and Tools

Programming Languages and Libraries: Python (Pandas, Numpy, Flask, Matplotlib), PySpark, Hadoop, Databricks, SQL

Machine Learning: Tensorflow, PyTorch, sci-kit learn, Transformers, feedforward neural networks, convolutional neural networks, feature engineering, transfer learning, natural language processing.

Experience

IBM, Cambridge, MA — 2022 – Present
Technical Account Manager (Data Scientist)

- Led the data science effort on an interdisciplinary team to generate a solution for a major client.
  - Working with limited compute resources, recommended and implemented a creative solution for outlier detection with over 1 million heavily skewed records.
  - After understanding the client's limitations and business context, explored multiple outlier detection approaches such as modified z-score, semi inter quartile range, and the medcouple algorithm.
  - Proposed the solution to the client, which walked them through all necessary steps and potential setbacks of the implementation, such as computational limitations.
  - Extracted and transformed data from client databases.
  - Due to scaling considerations, created a python implementation to conduct outlier detection in batches.
  - Coordinated progress with project management and automation experts.
  - Communicated and presented results to major client and internal stakeholders.
  - Results are currently being used to drive renewal negotiations with the client.
- Developing assets to improve client experience and internal operations.
  - Leading the development of an internal flask application to analyze and predict project health.
  - Working with internal leadership to understand their business needs to advance client negotiations.
  - Results are driving contract negotiations by worldwide sellers and solution architects.
  - Played a pivotal role in closing 4 renewal contracts for a large client.

University of California, Berkeley — 2022 – Present
Teaching Assistant (Machine Learning at Scale)

- Assisting in developing the future of the course while refining current material.
- Prepared a modeling dataset of over 1 billion records (1 TB of data) by cleaning, feature engineering, and joining tables.
- Providing guidance and feedback to current students with machine learning concepts.

FM Global, Johnston, RI — 2021
Technology Co-op II (Data Engineer)

- Created and deployed SQL views to production.
- Deployed SQL Server Integration Services (SSIS) packages to production using ETL design.
- Worked with data scientists to optimize their insurance metrics python script.
- Documented the team's data marts for retirement, in addition to creating a dimensional model.

Bryant University, Smithfield, RI — 2020-2021
Data Science Research Intern

- Developed modeling technique predicting hit songs to add to the 'Hit Song Science' literature.
- Fast tracked for journal publication in Volume 22 of the Issues in Information Systems (IIS).
- Developed a related teaching case published in Volume 20 of Information Systems Education Journal

Zachary Galante

Bryant University Football, Smithfield, RI 2019-2021

Head of Football Analytics

- Worked closely with the team's coaching staff to discover the most advantageous insights to be used in game.
- Effectively communicated modeling techniques and results to a non-technical audience.
- Using the flask microservice architecture, created a web-based application to predict a play's outcome in real time.
- Developed a machine learning pipeline to gather data used for real time evaluation.
- Results were submitted to the 2021 Horizon Conference with the San Francisco 49ers.

Massachusetts Parole Board, Natick, MA 2019- 2020

Research and Planning Intern

- Worked on a project focused on predicting recidivism in the state.
- Examined criminal histories containing confidential information to provide a clean dataset used for modeling.
- Results continue to be used by the Parole Board during release hearings.

Personal Projects

Brain Tumor Detection – UC Berkeley

- Worked on a team using over 4,500 MRI images to predict if a scan contains a tumor.
- Using TensorFlow, created a multilayer convolutional neural network achieving a recall of 99%
- Implemented a hashing function to deduplicate images and DB Scan to verify results.

NFL Big Data Bowl – Kaggle

- Used machine learning techniques to develop a performance metric to evaluate offensive line performance.
- Leveraged the results of a well performing model to create a transfer learning style approach model.
- Engineered innovative modeling features to account for data leakage.

Flight Delays – UC Berkeley

- Using map-reduce concepts, developed a modeling pipeline at scale to predict flight delays on over 42,000,000 records with over 200 features.
- Explored and implemented various approaches to handle heavy class imbalance.
- Engineered features, including personalized page rank, airline reputation score, and rolling averages to avoid data leakage.

Additional Machine Learning Projects

- Applied natural language processing concepts to implement synonym detection from scratch at scale.
- Using advanced map reduce skills, implemented a spam detection naïve bayes classifier at scale achieving an f-score of 0.88.
- With over 5 million Wikipedia pages, implemented a page rank graph from scratch.
- Used BERT and it's learned embeddings to predict and analyze IMDB movie reviews.