Caleb Tucker Dame

calebdame@gmail.com
www.calebdame.com
(707) 305-8453

Experience

Data Scientist – Transunion Feb. 2022 – Present
Remote Work

- Develop implementations of custom fraud detection models on client data using LightGBM/XGBoost, maintain a Python package for measuring drift in unlabeled data, and conduct research on fraud uses of Autoencoders, Stacked Models for predicting immature outcomes, and Graph Neural Networks.
- Lead cross-company collaboration initiatives and integration of newly acquired teams and data assets, resulting in operational efficiency, more powerful engineered features, and performance lift.

Data Engineer – M Science LLC Jun. 2021 – Mar 2022
Murray, UT

- Designed ETL frameworks and maintained 15+ pipelines that ran daily to export 3.5 Tb of custom data transforms to Financial Analysts using Databricks, Airflow, and Snowflake and worked with teams to perform quality assurance on existing pipeline logic and tested the purity and reliability of their data.
- Trained and parametrized clustering algorithms to regularly and automatically identify and flag potential noise and anomalies, increasing data purity by 10-15% on average. Changes were automatically visualized in Tableau dashboards to explore noise and anomalies to help explain changes to analysts and management.

Summer Data Analyst – Goldman Sachs Consumer Investment Management Jun. 2020 – Aug. 2020
Salt Lake City, UT

- Oversaw and designed an implementation of a new data pipeline that drew from databases across divisions using SQL and Alteryx to dynamically generate 30+ different SEC-compliant reports.
- Wrote scripts to find and extract necessary data from existing external reports to automatically store in spreadsheet format, saving 60+ hours monthly for the team.

Machine Learning Researcher – Brigham Young University, Economics Apr. 2020 – Jul. 2020
Provo, UT

- Trained ML models (XGBOOST, Neural Network, Naïve Bayes, etc.) in Python and managed a large database in SQL Server with the goal of linking census records across time and match duplicate genealogical profiles in the FamilySearch and Ancestry databases.
- Improved accuracy by 80% in linking census profiles by implementing advanced feature engineering informed by graph theory and increased efficiency by writing a custom implementation of the Hungarian Algorithm to finalize candidate matches when multiple classifications cannot coexist.

Education

Johns Hopkins University – M.S. Artificial Intelligence Aug 2022 – Present
Brigham Young University – B.S. Applied and Computational Mathematics & Economics Aug 2017 – April 2021

Languages and Competencies

Python Packages

- Numpy, Scipy, Matplotlib, Scikit-Learn, Keras, Tensorflow, Pytorch, StatsModels, Pandas, PyMC3, PyQT5

Other Languages, Software, and Frameworks

- SQL, Tableau, Pyspark, Unix Shell, Airflow, Snowflake, Databricks, AWS, C++, STATA, Git, and Agile

Projects

Battleship Solver

- Built battleship solvers with Neural networks. A Supervised model was built using labels from conducting Bayesian Sampling, and another model was built via Reinforcement Learning.

LSTM Mozart Music Generator

- Trained an LSTM Network to predict upcoming notes in Mozart Sonatas with accuracy of 50% when simultaneously predicting the correct next note's frequency, length, and offset. The Music21 package in Python was used to retrieve notes, chords, lengths, and note offsets from Mozart Sonatas formatted in .mid files.