

## Isaac McKillen-Godfried

[igodfried@isaac26.com](mailto:igodfried@isaac26.com) | ###.###.### | [www.linkedin.com/in/isaac-godfried-70874466](https://www.linkedin.com/in/isaac-godfried-70874466) | [www.isaacmg.github.io](https://www.isaacmg.github.io)

### Professional Experience

**Hybrid Data Engineer/Scientist** Monster Sept-2019 through Jul-2020

- Set company-wide standards for data science reproducibility best practices and trained coworkers in PyTorch.
- Utilized Terraform, GCS, BigQuery, and Pub/Sub to create data lake that both DS and BI teams leveraged.
- Trained and fine-tuned models in PyTorch to automatically add new triplets to the company knowledge graph in order to enhance down stream applications such as job search, candidate-job matching, and ad placement.

**Data Engineer** Hudson's Bay Company (Contract) Jan-2019 through May-2019

- Developed and optimized machine learning pipelines to forecast in-store retail demand with Spark MLlib library and algorithms such as XGBoost and linear regression. Utilized predictions for restocking optimization.
- Researched NLP techniques in PyTorch/Tensorflow to improve product categorization and personalization.

**Data Analyst** Eastern Maine Medical Center Jun-2016 through Jun-2018

- Refactored and implemented RNNs in Keras/Tensorflow to forecast patient length of stay and researched few-shot detection of conditions in Chest X-Rays with RetinaNet and YOLO2 to aid in clinical decision support.
- Created interactive visualizations for doctors and administrators with Bokeh, Pandas, and Jupyter Notebooks.

**Founder/Data Scientist** PaddleSoft Jun-2015 through Jun-2017

- Created a neural network (NARX) in MATLAB to predict the flow of the Kenduskeag stream.
- Employed Spark to perform analysis and train NLP algorithms like Word2Vec and LDA on textual datasets.

**Research Assistant** University of Maine Jun-2012 through Aug-2013

### Volunteer Work and Open Source Contributions

**Lead Machine Learning Researcher** CoronaWhy March-2020 to present

- Trained a relation classification model to automatically classify drug treatment pairs and adverse events from COVID-19 research articles, improved precision to 0.82 and recall to 0.88, and created Docker container to ship model to production in order to allow doctors to make better decisions about COVID treatments.
- [Leading a cross-disciplinary team](#) of data scientists, epidemiologists, and software engineers to utilize deep learning models (LSTMs/transformers) and transfer learning to forecast COVID-19 spread at county levels.

**Flow Forecast** Open Source Project August-2019 to present

- Developed transformer, LSTM, and regression models to forecast river and stream flows around the country.
- Created a deep learning for time series forecasting library to effectively track hyper-parameters/experiments. Library has now been adopted by several research groups and companies for its ease of use and accuracy.

### Technical Skills and Formal Education

- **Languages:** Python (PyTorch, Tensorflow, Keras, Bokeh, Pandas, and scikit-learn), SQL, and Scala
- **Technologies/Platforms:** Docker, Airflow, Kubernetes, Terraform, BigQuery, Pub/Sub, GCP, Wandb
- **Specialties:** Transfer learning, attention, transformers (HuggingFace), NLP (Spacy), Time Series Forecasting
- **Education:** B.A. Brandeis University Aug-2013 to May-2017

