Samuel Miller

samuelmillerwriting@gmail.com linkedin.com/in/samuelmillerwriting github.com/samuelmiller413

SUMMARY

Detail-oriented data scientist with the proven ability to work effectively with complex, big data from end-to-end through designing experiments, implementing state-of-the-art machine learning technologies, and generating data visualizations to present insight to non-technical stakeholders for increased access to data-driven-decision-making in an organization. Investing in my coding acumen through a 400+ hour coding bootcamp, I earned certificates in: SQL + Python Web Development and Data Science + Machine Learning, applying quantitative skills such as mathematics and computational statistics to perform regression and classification analyses though data collection, data cleaning, exploratory data analysis, and feature engineering. Demonstrating this stack of skills, I designed an experiment to predict solar energy potential by employing models like random forest ensembles, stochastic gradient descent, and a multilayer perception to deliver improved results over previous studies. As a dedicated learner, I am currently developing my passion for natural language processing and recurrent neural networks by studying deep learning under the mentorship of a doctoral-level industry professional. Combining this technical training with a degree in creative writing (Portland State University, Summa Cum Laude) and 3+ years experience as a freelance writer, I am eager to bring my expertise in written communication, collaborative design, teamwork, and storytelling to positively impact and innovate the way people interact with data and AI.

SKILLS

- **Languages**: Python, SQL, HTML
- **Libraries**: PyTorch, Lightning, Tensorflow, Keras, Sci-kit Learn, Numpy, Pandas, Matplotlib, Seaborn, SQL Alchemy, Gradio
- **Machine Learning**: data collection, analysis, visualization, feature engineering, clustering, regression, classification, decision trees**,** ensembles, SVM, PCA
- **Deep Learning**: language analysis, image classification, RNNs, CNNs**,** cloud GPUs, hyper-parameter optimization
- **NLP**: summarization, prediction, tokenization, sentiment analysis, BERT, ROUGE, transformers, TF-IDF
- **Platforms**: HuggingFace, Jupyter, Colab

In a data scientist role at your company I will…

- Collaborate with the data science team to design and conduct experiments to test models and glean insight from data.
- Integrate my training as a creative writer to construct meaningful narratives from the data, generating incisive data visualizations to offer impactful data storytelling to stakeholders.
- Leverage my written and verbal skills to communicate technical concepts to non-technical audiences.
- Ensure the quality of data throughout the entire pipeline by collecting, cleaning, and structuring data with best practices.
- Work with teams from other departments to holistically devise data-driven solutions on time for multiple projects.
- Cultivate inclusivity, teamwork, innovation, mentorship, organization, and mutual support in the workplace.
- Diligently prioritize multiple projects to ensure timely results.
- Be coachable and accepting of constructive criticism.

PROJECTS

Predicting Solar Energy **Git:** github.com/SamuelMiller413/Predicting-Solar-Energy
Tech Talk: youtube.com/watch?v=aFapydKwPcM&feature=youtu.be

Designed an experiment to predict the solar energy potential of a given region using meteorological data and to study the role that recursive feature elimination serves in working with the data. built several machine learning models, yielding improved accuracy over models of similar studies through the help of feature engineering.

- Conducted extensive data analysis, domain research, and feature engineering to test hypothesis
- Engineered Random Forest Regressor, Stochastic Gradient Descent, and Multi-layer Perceptron
- Employed cross-validation and metrics such as mean squared error, R2, and others to quantify error.
- Worked with a senior data scientist mentor, eventually publishing the paper and presenting it as a Tech Talk.

Text Summarizer **HuggingFace:** huggingface.co/spaces/SamuelMiller/sum_it

Built an NLP application to take in text from users and summarize it using text extraction. Utilized transfer learning to develop the model from Google's T5 Transformer, implementing the Gradio library and building the web application.

- Built and trained deep learning model using colaboratory's TPU.
- Experimented with hyper-parameters, masking, and various optimizers to improve performance.
- Evaluated model performance using Rouge and BERT metrics.
- Delivered Model in the form of a web application using Gradio.

EDUCATION

CodingNomads - **Certificates: Python Web Development + SQL** **Online** 2021
Data Science + Machine Learning **Online** 2022
Deep Learning (in progress) **Online** 2022

Learned end-to-end machine learning techniques including data collection, data analysis, feature engineering, experiment design, model building, testing, and presenting results through data visualization and writing. Regression and classification analyses through decision trees, gradient descent, and clustering. Trained with tools such as Sci-kit Learn, Matplotlib, Numpy, Pandas, and Seaborn. Trained on software engineering best practices and fundamentals of Python programming such as data structures, algorithms, networking, unit testing, and version control. Worked with relational databases and developed REST APIs using SQL and SQL Alchemy.

Portland State University - Bachelor of Fine Arts in Creative Writing: Fiction **Portland, OR** 2017-2020

EXPERIENCE

Freelance Writer **Portland, OR; Santa Cruz, CA** 2019 - Present
Articles: “How to Learn Python: A Beginner’s Guide,” “What Does a Data Scientist do?” (CodingNomads Blog)