BAILEY WEI

New York, NY | 646-371-6126 | Work Authorization: U.S. Citizen
bw489@cornell.edu | linkedin.com/in/baileywei | github.com/bw-tech | baileywei.com

EDUCATION

Cornell University, New York, NY — May 2020
Master of Engineering in Operations Research and Information Engineering – Data Science Concentration
Honors/Awards: Cornell Merit Scholarship
Relevant Coursework: Deep Learning, Natural Language Processing, Data Science in the Wild, Applied Machine Learning, Optimization Methods, Modeling Under Uncertainty

University at Buffalo, Buffalo, NY — May 2019
Bachelor of Science in Industrial Engineering
Honors/Awards: Cum Laude, Dean's List

TECHNICAL SKILLS

Programming Languages:	Python, R, SQL, Swift, C++, MATLAB
Frameworks and Libraries:	Matplotlib, NLTK, Pandas, PySpark, PyTorch, Scikit-Learn, Spark, Transformers
Operations Research:	Gurobi Optimizer, Six Sigma Green Belt
Additional Tools:	AWS, Azure, CUDA, Google Colab, Tableau

EXPERIENCE

Cornell Tech, *iOS Engineering Intern*, New York, NY — Dec. 2019 – Present
- Developed an iOS application using Microsoft Computer Vision API to generate captions for taken photos
- Integrated Amazon S3 to create data storage solution for real-time result and feedback collection
- Enhanced accessibility for vision-impaired users by implementing VoiceOver and other audio cues

Nielsen, *Data Science Intern*, Columbia, MD — Jun. 2019 – Aug. 2019
- Utilized U.S. Census and internal data to develop random forest, k-NN, and logistic regression models to identify hard-to-survey households
- Increased F1 score by 150% and reduced household false positive rate by 50% over existing models
- Enhanced data preprocessing methodology through reducing discarded radio surveys to save $300K annually
- Promoted continuous learning and integration of relevant industry practices by leading weekly forums for a team of 11 scientists

Sentient Science, *Predictive Analytic Modeling Intern*, Buffalo, NY — Feb. 2019 – May 2019
- Developed a visualization library in Python to automate graph creation, reducing deliver times of KPIs to leadership by 10%
- Validated and predicted wind turbine failures by creating regression models to support data scientists
- Optimized data aggregation from Amazon S3 by reducing data acquisition time by 15% per visualization

PROJECTS

Fine-Grained Sentiment Classifier, arXiv:2005.13619, (Google Colab, Python, PyTorch) — Spring 2020
Created a classifier to predict sentiment ratings on the Stanford Sentiment Treebank dataset
- Evaluated several BERT-like models to determine accuracy and complexity trade-offs
- Utilized Google Colab GPUs to establish state-of-the-art accuracy, 60.2%, using RoBERTa-Large models

Fake Twitter News Detection, (AWS, NLTK, Python, PyTorch, Scikit-Learn) — Spring 2020
Identified fake news related to politics and the U.S. election through deep learning solutions
- Developed pipeline to extract and preprocess a dataset of over 6,000,000 unique tweets
- Utilized AWS along with NLTK and PyTorch to create BERT embeddings which reached 99.81% accuracy

Cornell Tech, BigCo Studio Challenge: Intersection, (HTML, JavaScript, Python, SQL) — Spring 2020
Developed a trivia web application to integrate with Intersection kiosks to push user engagement and collect additional data
- Analyzed various product ideas to best facilitate kiosk-to-phone experience and increase kiosk attractiveness
- Validated long-term user retention through conducting several experiments on live Intersection kiosks

Cornell Tech, Product Studio Challenge: JPMorgan & Co. (Python, Scikit-Learn, Swift, Tableau) — Fall 2019
Developed a recommender system to predict changing neighborhoods and demographics
- Utilized Instagram, Twitter, and Yelp data to further push accuracy of neural network models
- Built iOS application to implement recommender system and visualize long-term neighborhood shifts

Large-Scale Image Search Engine, (Python, Scikit-Learn) — Fall 2019
Created a model to retrieve relevant images based on text inputs
- Utilized TF-IDF, Word2Vec, ridge regression, ResNet, and neural networks to classify 85% of correct images
- Developed color-based feature vectors for additional data to drive predictive power and accuracy