#### Frank Wei, Ph.D.

(385)394-2509, [fweital@gmail.com](mailto:fweital@gmail.com),

Austin, TX

8+ years industry and research experiences in Machine Learning/Deep Learning/Statistical experiment design from POC development to large scale product implementation (deploy). 5+ years experiences as tech lead.

# Skill set

Programming languages and packages: Python, R, Tensorflow, Pyspark, Pytorch, SAS, SQL, Matlab, Keras, OpenCV, NLTK

Technique skills: Machine learning/Deep learning/Predictive modeling, Computer Vision (CV), Natural Language Processing (NLP), AWS

#### Professional Experience

**General Motors, Austin, Texas, 2019 - Now**

**Lead AI/ML Scientist**

* Lead a team of scientists for High Definition Map roadway creation in GM’s L3 autonomous driving system – Ultracruise
  + Participate in planning, roadmap, and architecture discussions to help evolve the algorithms into software. Merge roadway features with the result from the map intersection creation team using intersection tie-in points, and delivered the phase one result.
  + Based on the high-resolution Aerial images, create roads and lane edges using segmentation deep learning models (deeplabv3); conduct road type classification, lane marker type classification, and road edge type classification using Inception-ResNet; Object detection on-road objects (stop bar, crossings, on-road text) using FAST-RCNN.
  + Based on HSVT data, create road centerline using spline in the production phase; using additional high-resolution Aerial images, an encoder (ResNet) -decoder (LSTM) with attention mechanism is used to generate centerline points, and the process is still in the research phase; All the processes are based on python, pytorch, and OpenCV.
  + Construct a pytorch pipeline for model deployment, and deploy road features based on Maglev, an Nvidia computational platform, through a ymal file using docker image in the artifactory for the large scale implementation, and output a workflow that contains XML format result with Edge Id /Crossing Id /Latitude /Longitude /Altitude /road feature attributes.
  + Mentor a team of developers for the map quality check using sanity check, and ground truth Lidar data based on Pyspark. The result is visualized and delivered to our clients.
* Lead electronic vehicle infrastructure
  + Participate in planning, roadmap, and architecture discussions to help evolve the algorithms into software that display the existing potential EV charging stations. Meet with inside and outside investors and help map business needs into product requirements.
* Use Pyspark to pull the high-speed vehicle telemetry (HSVT) time-series data from Hive, and process the data using trip merger, and charge event extractor modules to the vehicle-trip level attribute-relation table, and upload the table in S3 bucket for the feature engineering.
* Predict charge events based on engineered features and past charge events using the LSTM model. Identify the optimum location for charging infrastructure using DBSCAN based on the existed and predicted charge events. Using K-Means clustering to identify the potential electronic car buyers from the internal combustion engine driver.
* Deploy the clustering model using flask for our client to visualize.
* Mentor junior AI/ML scientist, and conduct model and code review for peers. In the process of submitting several papers.

**National Institute of Statistical Sciences & U.S. Department of Agriculture, Washington D.C., 2013-2019**

**Senior Research Statistician**

Help USDA implement AI solutions to different scenarios.

* Led hog/pig official estimates prediction team.
* Used SQL to pull data and used python to merge, impute, and process hog/pig survey data and hog/pig official estimates for ratio adjustment in addition to calibrating state weights. Updated the previous model using a new SARIMA model to generate quarterly publish estimates for hog/pig national numbers, resulting in improving the performance by 50%.
* Expanded the SARIMA model to finer temporal (quarter to month) and spatial (national to state) using survey data. Used web-scraping and natural language processing (NLP) in Pytorch and NLTK to detect hog/pig disease outbreak (shocks) in the United States and extracted outbreak information.
* Constructed an artificial intelligence system that processing disease outbreak detection, information extraction simultaneously using TFIDF, logistic regression for intent classification, BERT for Named Entity Recognition (NER), and information extraction (IE) using Pytorch.
* Led a team to automate book classification by multimodal approaches
  + Collaborated with DC public library, classify books into different types based on their preface, introduction, and the cover.
  + Used TFIDF, LDA, w2v, n-grams, and ResNet for feature engineering from text and image, and a LightGBM is used for the classification. The model was deployed using flask.
* Derived text classification, sentense similarity, sentiment analysis for the survey answers using LSTM and w2v based on NLTK, Spacy, and Pytorch. The model was deployed using AWS SageMaker.
* Developed an artificial intelligence system in Tensorflow that synthesizes two million records census of agriculture dataset using Generative Adversarial Networks (GAN) to remove the personal identifiable information (PII).
* Designed a unified deep learning model for imputation, dual system estimation, and calibration simultaneously for the census of agriculture by python and Tensorflow in AWS.
* Create the Cropland Data Layer (CDL) for the entire United States using CART based on over 100 bands satellite images to classify crop and yield types.
* Evaluated different designed-based (PRNs, CRNs), and model-based approaches to reduce respondent burden, and the result outperforms 20% of the current used approach SIP on Crops Acreage, Production and Stocks (Crop APS), Agricultural Yield Row Crops (AYS-RC) and Agricultural Yield Small Grains (AYS-SG) surveys.
* Processed, imputed, edited, and merged more than ten data sources for 2 million records census of agriculture, and prepared predictive modeling using Pyspark. Developed a logistic regression model to predict national farms that are not on the list (NML) to prepare calibration. Allocate national farms for 2017 census of agriculture by optimization.

##### Selected Publications

* (**Under review**) Wei, Y, and R. Yang, 2020: Using Multimodal Datasets to Automatically Identify Important Features in the Tropical Cyclone Rapid Intensification Using An Advanced Artificial Intelligence System. *Weather and Forecast*
* (**Under review**) Wei, Y, F. Mahnaz, Y. Mengistu, S. Mahesh, and O. Bulan, 2020: Creating HD map using aerial Images and Vehicle telemetry data. *IEEE Transactions on Intelligent Transportation Systems*
* (**Under review**) Wei, Y, Q. Zhang, Y. Mengistu, S. Mahesh, O. Bulan, and F. Mahnaz, 2020: Road Categorization Using Aerial Imagery and Vehicle Telemetry for Creating HD Maps for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems*
* (**Under review**) Wei, Y and R. Yang, 2020: An Advanced Artificial Intelligence System for Investigating the Tropical Cyclone Rapid Intensification with the SHIPS Database”. *Weather and Forecast*
* (**Under review**) Wei, Y, N. Sedransk, and L. Sartore, 2020: Web Scraping and Natural language processing for disease outbreak detection and information extraction. *Applied Stochastic Models in Business and Industry.*
* Sartore, L., Y. Wei, E. Abayomi, S. Riggins, G. Corral, V. Bejleri, C. Spiegelman, 2019: Modeling swine population dynamics at a finer temporal resolution. *Applied Stochastic Models in Business and Industry.*
* Corral, G., S. Riggins, E. Abayomi, L. Sartore, Y. Wei, N. Sedransk, C. Spiegelman, and L. J. Young, 2019: On Producing Estimates for NASS’s Quarterly Hogs and Pigs Report. *2019 National Agricultural Statistic Service*, USDA, Washington, DC.
* Wei, Y., L. Sartore, J. Abernethy, D. Miller, and K. Toppin, 2018: Deep Learning for Data Imputation and Calibration Weighting. In *Proceedings of the 2018 Joint Statistical Meetings*, Vancouver, Canada.
* Wei, Y. and V. Bejleri, 2017: Exploring Sampling Techniques to Reduce Respondent Burden. In *Proceedings of the 2017 Joint Statistical Meetings*. Baltimore, MD.
* Wei, Y., 2016: Agricultural Coverage Evaluation Survey (ACES) documentation. *2016 National Agricultural Statistic Service*, USDA, Washington, DC.

##### Education

**Ph.D. in Computational Science and Informatics,** 2016-2020, George Mason University, Falls Church, VA

**Master of Science in Applied Statistics,** 2012 - 2013, University of Michigan, Ann Arbor, MI

**Master of Arts in Applied Mathematics,** 2011 - 2012, University of Michigan, Ann Arbor, MI