Fred Mubang, PhD

Email: fmubang@usf.edu Machine Learning Researcher / Data Scientist Linkedin: Fred Mubang
Website: fmubang.github.io

Summary: PhD Machine Learning Researcher and Data Scientist with 5 years of work experience. I am currently a Senior Data Scientist at Experian. Before that, I spent most of that time working for the Defense Advanced Research Projects Agency (DARPA), an agency within the Department of Defense. Also during that time, I was working on my Ph.D. in Computer Science from the University of South Florida, which I completed in October 2022.

Work Experience

Senior Data Scientist at Experian **Oct 2022 - Present**

- Performed data analysis and build AI/machine learning models on marketing and credit data for hundreds of millions of individuals.
- Presented insights to upper management and business stakeholders in the form of various Power Point presentations and reports.
- Saved organization $20+ million in potential lost revenue due to usefulness of data analysis and models built.
- Heavily utilized Python, Pyspark, Hadoop, Pandas, Scikit, and Numpy for data analysis and model building.

Machine Learning Researcher/Data Scientist for DoD/DARPA Social Simulation Project - Link **Oct 2017 — Oct 2022 (5 yrs.)**

- Objective: High-fidelity computational simulation and of user activity across different social media platforms such as Twitter and YouTube, among others.
- Built neural network and XGBoost machine learning models to perform time series forecasting and network simulation of user activity in various social media platforms with millions of users. Achieved over 20% improvement over historical baselines in both time series and network simulation tasks.
- Performed various data engineering tasks such as cleaning, manipulating, scraping, feature engineering, and visualization of data.
- Performed detailed social network time series analysis of various datasets and created weekly Powerpoint presentations containing data analysis and insights. Used various Python libraries to prepare results such as Scikit, Pandas, Matplotlib, and Numpy.

AI Bootcamp Instructor **May 2020 — Aug 2020**

- Prepared coursework and taught students various Artificial Intelligence concepts

Education

Ph.D in Computer Science, University of South Florida, GPA: 3.87/4.00 **Aug 2018 — Oct 2022**
Master of Science, Computer Science, University of South Florida, GPA:3.87/4.00 **Aug 2018 — May 2021**
Post Bachelor Studies, Computer Science, University of South Florida, GPA:3.7/4.00 **Aug 2017 — July 2018**
Bachelor of Arts, Music Business, Berklee College of Music **Aug 2010 — May 2014**

Relevant Courses: Data Mining, Machine Learning, Neural Networks, Advanced Neural Networks, Social Media Mining, Network Science, Natural Language Processing, Intro to AI, Calculus 1-3, Linear Algebra, Probability and Statistics

Skills and Technologies

- **General ML Skills:** Classification, Regression, Clustering, NLP, Feature engineering, Dimension reduction techniques, Logistic Regression, Linear Regression
- **Data Analytics:** Descriptive Statistics, Cleaning, Manipulation, Scraping, Visualization
- **Technologies:** Python, C, C++, Linux, Scikit-learn, Pandas, Pyspark, Hadoop, Tensorflow, Keras, XGBoost, Numpy, Seaborn, Matplotlib, Networkx, Jupyter Notebooks, Excel

Publications (With Links)

- **Mubang, F.**, *Social Media Time Series Forecasting and User-Level Activity Prediction with Gradient Boosting, Deep Learning, and Data Augmentation.* Ph.D dissertation, College of Computer Science and Engineering, University of South Florida, USA, 2022. - Link
- **Mubang, F.**, Hall, L.O. *VAM: An End-to-End Simulator for Time Series Regression and Temporal Link Prediction in Social Media Networks. IEEE Transactions on Social Computing* (2022 - In Press) - Link
- **Mubang, F.**, Hall, L.O. *Simulating User-Level Twitter Activity with XGBoost and Probabilistic Hybrid Models. arXiv preprint arXiv:2202.08964* (2022) - Link
- **Mubang, F.**, Hall, L.O. *A Survey of Recent Artificial-Intelligence Driven Frameworks for User-Level Activity Prediction in Github. Engineering Applications of Artificial Intelligence* (2022 - Under Review) - Link
- Liu, R., **Mubang, F.** *Simulating Temporal User Activity on Social Networks with Sequence to Sequence Neural Models. IEEE SMC International Conference* (2020) - Link
- Liu, R., **Mubang, F.**, et al. *Predicting Longitudinal User Activity at Fine Time Granularity in Online Collaborative Platforms. IEEE SMC International Conference* (2019) - Link