# Adam C. Yang

7911 Westpark Dr. Apt. 524, Mclean, VA 22102 · (217)550 –2667 · adamcy99@gmail.com
https://www.linkedin.com/in/adam-yang/

Electrical Engineer and Data Scientist experienced with harnessing data in the development of complex, cutting edge, semiconductor microprocessors. Pioneered and led the integration of machine learning and data science concepts within my IBM team to address engineering challenges and eager to expand my capabilities of extracting value from data.

## Data Science and Machine Learning Experience

### IBM

*Hardware Developer, Advisory Engineer* April 2020 - Present

- Engineered a robust big data pipeline using Python and SQL, incorporating an algorithm to analyze 9,000+ unique parameters, with the objective of isolating significant shifts. This automated the timely detection of potentially harmful changes in electrical or device measurements, most of which are not regularly monitored, resulting in saving IBM $100,000s in potential costs.
- Developed and maintained seven dashboards using Python and SQL. Automated engineering tasks and introduced innovative data visualization methods through a robust user-interface. Leveraged data science libraries such as Scikit-Learn and Dash. Resulted in 10s of hours of reduced engineering workloads and facilitated in uncovering novel insights.
- Primary owner and lead of SRAM and Logic yield for 7nm and 5nm IBM microprocessor products. Implemented machine learning techniques to classify multiple systematic defects, which resulted in a 20% yield improvement and maintained a baseline of 98% yield. Utilized Random Forest Classifiers and custom-built classifiers to classify defects.

*Product Yield and Characterization Engineer* March 2016 - April 2020

- Oversaw SRAM and Logic related defect monitoring and analysis for 14nm IBM microprocessor products. This included extensive data mining for correlations and trend analysis to isolate device flaws and diagnosis of systematic and random defects. Coordinated daily with multiple engineering teams. Maintained an SRAM and Logic yield of over 90%.
- Pioneered the implementation of Random Forest Classifiers and custom-built classifiers to identify and monitor up to 10 unique random and systematic defects, resulting in multiple instances of SRAM and Logic yield improvement. Clearing up these defects resulted in yield enhancements as high as 60%.
- Analyzed the impacts of device gate height to yield and metrology parameters to help evaluate whether the increased performance justifies the potential yield loss. This directly impacts the business decision of finding the correct balance between cost and performance.

## Education

**University of California, Berkeley** Graduated: Dec 2019
Master of Information and Data Science

**Georgia Institute of Technology** Graduated: Dec 2015
Master of Science, **Electrical Engineering**
Minor in **Computer Science**

**University of Illinois at Urbana-Champaign** Graduated: May 2014
Bachelor of Science, **Electrical Engineering** - Graduated with Honors/Dean's List

## Relevant Academic Experience

- **Experiments and Causality –** Designed and conducted a controlled field experiment to determine the effects of political news articles on stress.
  ***Topics*:** R, experimental design, statistical analysis, causal inference, cleaning data, A/B Testing
- **Machine Learning at Scale –** Implemented word count on a large text and composed a Naïve Bayes algorithm to classify spam mail using Hadoop MapReduce. Established an algorithm to detect synonyms using Spark and a MapReduce framework, applied gradient descent to supervised learning algorithm to predict wine quality, and generated a PageRank algorithm. The final project involves predicting advertisement click-through-rate based on a Kaggle competition by Criteo.
  ***Topics:*** Apache Spark, Hadoop MapReduce, Google Cloud Platform (GCP), parallel and distributed computing.
- **Natural Language Processing** – Created custom word embeddings based on raw text data from Reddit. Then built a sentiment model based on biased political embeddings to analyze each group's difference of opinion on certain topics and people.
  ***Topics:*** BERT, ELMo, Neural network models, CNN, RNN, LSTM, language models, word embeddings, sentiment analysis, machine translation and summarization, part of speech tagging

## Skills

***Languages:*** Python, SQL, R, C
***Libraries*:** Scikit-Learn, NumPy, Pandas, Maplotlib, TensorFlow, Dash, Seaborn, PySpark
***Miscellaneous:*** supervised and unsupervised machine learning, A/B Testing, Google Cloud Platform (GCP), Amazon Web Services (AWS), Hadoop MapReduce, Spark, JMP, Git

**U.S. Patent 20190245032,** "Vertical Memory Cell with Mechanical Structural Reinforcement" Jan 5, 2021