Samba Njie Jr.

Foster City, CA | 408-250-8332 | njie.samba209@gmail.com | linkedin.com/in/sambanjie | snjie209.github.io

Education

Stanford, Graduate Certificate (Courses in Artificial Intelligence/Statistics), 2019-2021

UC Berkeley, B.A. *Statistics*, B.A. *Applied Mathematics*, August 2017

College of San Mateo, A.S. Computer Science, Expected 2021

De Anza College, A.S. Biological Sciences, Magna Cum Laude, Honors Scholar, Phi Theta Kappa, June 2014

Coursework: ANOVA/Regression Models, Deep Learning, Machine Learning, Data Science & Visualization, OOP in Python, Numerical Analysis, Time Series, Optimization Models in Engineering, Probability, Mathematical Statistics & Data Analysis, Linear Algebra, Object Oriented Programming in Java, Data Structures, Enterprise Database Management

Skills

Languages: Python, Oracle SQL, MySQL, Linux/Shell Scripting, Java, R, Matlab, XML, Javascript, MS Office

Technical: Git, Docker, Jupyter, Scikit-learn, Pandas, Numpy, SciPy, TensorFlow, OpenCV, LaTeX, ggplot2

Professional Experience

Data Scientist II, Oracle, Redwood Shores, CA   *June 2018 - Present*

- Productized 2 enterprise AI products with $500k+ ARR (customer churn, engagement scoring, lookalike modeling)
- Completed data science projects from product requirements assessment, MVP scoping, research, implementation, testing, and deployment (using Docker, Jenkins, unit testing, etc.)
- Employed reinforcement learning (MAB, MDP, contextual bandits), clustering analysis, statistical methods, etc. for proof-of-concept development of new features
- Created three pending patents that are productized in Responsys/Eloqua Marketing Cloud product environments
- Interfaced with cross-functional teams (product, UX, UI, backend engineering/ETL, management) to develop product strategy and decisions based on data analysis
- Organized and participated in hackathon for 200+ engineers for rapid idea generation, innovation, productization

Machine Learning Research Assistant, Oracle Labs, Redwood Shores, CA   *July 2017 – October 2017*

- Performed bound and bottleneck analysis on state-of-the-art algorithms such as convolutional neural networks and support vector machines, yielding 99%, 80% and 73% test set accuracy on MNIST image recognition classification.
- Developed Python based tools for profiling, including runtime and memory profiling that observes analysis within call-stack of each function, performance counters to identify hardware-specific bottlenecks of algorithms
- Assisted in instruction of 90% of Ph.D. and Masters level engineers about machine learning theory and optimized computation frameworks such as TensorFlow, yielding in almost 60% retention in usage of TensorFlow for team use.

Data Science Intern, Gronade, Sydney, NSW, Australia   *May 2017 – July 2017*

- Trained image classification models using contour pattern recognition to detect image labels for 95% test accuracy
- Worked with databases to develop time series analysis plots for use in online marketing platform

Data Analytics Research Assistant, UCSF Multiple Sclerosis Lab, San Francisco, CA   *September 2016 – September 2017*

- Prototyped and cleaned large-scale 27 million case data of patient records, optimizing implementation by 25%
- Designed over 10 data cleaning, data mining, extractions, and statistical analyses algorithms on patient data

Projects

- **Predicting Low Vision Patients Using Deep Neural Networks**   *June 2020*

  Preprocessed structured patient visit data and clinical notes text using NLTK, word2vec word embeddings, and used MLPs and autoencoder models to predict prognosis of patients with low visual acuity, achieving AUROC of 80%.

- **Hilary Clinton's Unclassified Emails: Machine Learning Prediction on Sender Classification**   *November 2016*

  Built classification algorithms such as random forests, trees, support vector machines, k-means clustering, and utilized natural language processing to parse documents and predict and classify common words for each sender.