Tatenda Ndambakuwa

tn1133@nyu.edu | 1-210-529-2591 | linkedin.com/in/tatenda-ndambakuwa | github.com/Tatenda-Ndambakuwa

EDUCATION

NEW YORK UNIVERSITY, New York, NY   May 2020
Masters of Science in Data Science | Capital One Fellow, AAUW International Fellow, Salesforce Quip Scholar

VIRGINIA COMMONWEALTH UNIVERSITY, Richmond, VA   May 2018
Bachelor of Science in Applied Mathematics | Minors: Urban Planning, Physics   GPA: 3.5/4.0

RELEVANT COURSEWORK

Deep Learning for Natural Language Processing, Deep Learning for Medicine, Machine Learning, Big Data, Recommender Systems, Probability & Statistics, Data Visualization, Inference and Representation, Computer Vision.

RELEVANT SKILLS

Languages: Python, R, Scala, Java, JavaScript.
Parallel Programming: Multithreaded/distributed computing: CUDA, OpenMP, and MPI.
Packages: ArcGIS, OpenCV, SciKit-Learn, PyTorch, TensorFlow, Keras, SciPy, Pandas, NLTK, spaCy, CoreNLP, Gensim, Word2vec, GloVe.
Big Data Stack: SQL, Hadoop, MapReduce, Spark, Docker, Google Cloud Platform, Amazon Web Services, Docker, Kubernetes, GPUs
Machine Learning/ Statistics: Linear/Logistic Regression, SVM, Random Forests, Decision trees, k-NN, Naive Bayes etc.
Visualization: Tableau, D3.js, Bokeh, Datashader, Holoviews, Plot.ly, ggplot, HTML5, CSS3.

PROFESSIONAL EXPERIENCE

New York University, New York, NY   August 2019 - Present
Graduate Researcher | Social Media and Political Participation (SMaPP) Lab.
- Used Twitter data to create the Joint Embedding of User-content and Network Structure to Enable a Common Coordinate that Captures Ideology, Geography and User Topic Spectrum.
- Created embedding using Modular Non-Negative Matrix Factorization(M-NMF) and Large-scale Information Embedding (LINE).

NVIDIA, Austin, TX   May-August 2019
Data Science Intern (AI Infrastructure) | Rapids.ai
- Implemented pandas, dask, spark and Python work flows, on GPU, using the Rapids framework.
- Benchmarked and profiled ETL processes and suggested improvements to document features.
- Created data visualizations, anomaly detection frameworks, machine learning models and predictive models for use cases.

USAID, Washington, D.C   September-December 2017
Data Science Intern | US Global Development Lab
- Conducted cost-benefit analyses and inclusive growth diagnostics for projects funded by USAID.
- Used R to perform various statistical techniques, including regression, chi-square, Kruskal- Wallis Rank test, cluster analysis, Pearson's correlation, Z-scoring, T-Testing and ANOVAs. Conducted data cleaning, data analysis and visualizations, and data validation using R.

AEM CORPORATION, Herndon, VA   June-August 2016/7
Data Science Intern | Software Consulting
- Used statistics to create default budgets for several age groups and income levels for a retirement financial application called plynty.
- Performed extraction, processing and cleaning using R for data from the US Bureau of Labor and Statistics for all 50 states.
- Utilized NLP approaches and Latent Dirichlet allocation to develop smart recommender systems for marketing.
- Used JavaScript, AngularJS, node.js, JSON, Ionic, and HTML5, CSS3, to create Google and Mapbox maps for use in a GIS-based online relocation recommendation tool for a retirement financial application.

SHIRI, Richmond, VA   August 2016 - Present
Founder & Developer | Agricultural Technology Startup
- 2019 Agri Innovator of the year, Women in Agriculture Zimbabwe
- Created a mobile application that establishes a new peer-to-peer communication platform for African farmers. The Shiri platform allows farmers to access resources, suppliers and vendors using their mobile phone.
- Developed UI/UX workflow and design using InDesign and Sketch. Conducted user research, partnership development and marketing.

PROJECTS

- **Natural Language Understanding: Identifying words/phrases varying by targeted gender, as utilized by ISIL on Twitter in reaching out to audience/potential recruits**

Trained word2vec vectors and performed linear mapping on 18 million tweets. Determined masculine and feminine word associations using ultra dense embeddings. Validated the word embeddings using word-similarity metrics.

- **Recommendations: Recommender System for Music Data**

Used Spark's alternating least squares (ALS) method to learn latent factor representations for users and items. Used the learned representation to develop a UMAP visualization of the items and users.

- **Natural Language Processing:** Transfer Learning, Question Answering, Neural Conversation Modeling (modeling, decoding, interactive chat), Machine Reading
- **Inference and Representation:** Implementing Hidden Markov Models for Speech Recognition