Gebreab K. Zewdie, PhD
Data Scientist

Cell Phone: 214-518-3266, Email: **phygbki21@gmail.com**,
https://www.linkedin.com/in/gzewdie/, https://github.com/gzewdie

SUMMARY

- Data Scientist with 8+ years of experience in big data, machine learning, deep learning, predictive modeling (time series forecasting), data mining, statistical and optimization methods. Strong programming skill in Python, Matlab and SQL, working knowledge in R, Scala and C++. Familiar with NLP, PySpark, cloud platforms (GCP and AWS).

- Self-motivated, with courage and ability to handle challenging projects while consistently meeting deadlines.

PROFESSIONAL EXPERIENCE

Verizon, *Data Scientist, Atlanta, GA*　　　Mar. 2021 - Present

Working with a dynamic team of data scientists to solve complex business problems at Verizon based on big customer data and advanced machine learning methods.

- Working on security analysis on datasets coming off kafka to classify events as malicious or not. Various events (emails, mesages, etc) are posted on kafka. I am working on pulling data from kafka and developing machine learning models to classify each event as malicious (spam) or genuine.

- Working on developing end-to-end machine learning models for financial forecasting in partnership with the financial team using big customer data sets to forecast various revenues and costs. Main tasks include pulling useful data in batch from Teradata in SQL query running from Python environment, pre-processing the data, testing, selecting, and tuning machine learning methods.

- Developed data-driven automated best-performance machine learning applications for forecasting various profitability and cost matrices 36 months ahead connecting to Teradata Database from Python on a Unix-based server.

- Working on improving machine learning models regularly, collaborating closely with engineers to deploy machine learning models in production in a batch process, and systematically tracking model performance.

- Participating in Verizon's survival model development and building Rapid Simulation Engine (RSE) to help stakeholders make important decisions.

Georgia Tech, *Data Scientist, Atlanta GA*　　　Aug. 2019 - Mar. 2021

Worked with a cross-functional team as a data scientist to develop machine learning based innovative solutions to space weather risks for early warning, and mitigation. My main responsibilities include:

- Identify, acquire and analyze large unstructured and (semi-) structured data sets, mine the data for pattern detection, engineer features, and select useful features for optimization, visualize the data. I use mainly Python, Scikit-learn, Tableau, PySpark.

- Implementing multivariate LSTM and fully connected neural network over TensorFlow using the Keras API for forecasting key space weather parameters a head of time.

- Used the AWS cloud computing platform for model deployment using docker container and model integration.

- Understand key business (client) requirements and develop data-driven solutions for early detection and recommendation.
- Participating in developing end-to-end machine learning pipeline to standardize processes, model integration and deployment on cloud.

UT Dallas, Quantitative Researcher (Data Scientist role)
Aug. 2013 - Jul. 2019

I worked on applying data science techniques to estimate allergic pollen and other environmentally hazardous particulates over the Dallas-Fort Worth area. My achievements are:

- Analyzed large structured and unstructured data sets, implemented machine learning to forecast the abundance of allergic pollen and other particulate matter using Python and R.
- Developed near real-time tracking of pollen allergy outbreaks using NLP techniques to social media data and determine its relation with wind pattern and pollen source.
- Consulted the Texas Commission on Environmental Quality to optimally invest on air quality monitoring network of sensors.
- Processed large radar array data, developed inversion techniques to these data to image the electron distribution and applied Self-Organizing Maps machine learning to identify interpretable patterns.

IGSSA, Addis Ababa University, Data Analyst   Jan. 2010- Jul. 2013

- Extracted, cleaned and analysed large GPS data to identify ionospheric TEC patterns and develop tomographic inversion techniques with L2 regularization in Matlab and Fortran.
- Estimated precipitable atmosphere water content from GPS signal delay.

SAMPLE PROJECTS

- https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020SW002639
- Machine Learning Project, Publication: *https://doi.org/10.3390/ijerph16111992*
- LSTM to Forecast Apple Stock: *https://github.com/gzewdie/Apple-Stock-Forecasting*

SKILLS

Machine Learning (8 years), Data Analysis (11 years), Statistical Analysis (12 years), Time Series Analysis (6 years), Deep Learning (4 years), Github (7 years) ,Data Visualization (12 years), Python (9 years), Numpy (8), Pandas (8 years), Anaconda(conda, mini Conda, 6 years) R (4 years), Scala (2 years), SQL (5 years), Jupyter notebook, JupyterLab, Gitlab,Teradata (2 year), Apache Spark (3 years), Scikit Learn (8 years), Hadoop (3 years), Pandas(8 years), Tableau(2 years), Matlab (12 years), Jenkins (2 years), TensorFlow (4 years), PyTorch (4 years), AWS and HPC (4 years), Docker (2 years).

EDUCATION

Ph.D in Physics (Advisor: Prof. David J. Lary)
The University of Texas at Dallas, Dallas, TX, USA
M.S in Space Physics
Addis Ababa University, Addis Ababa, Ethiopia
B.S in Physics
Dilla University, Dilla, Ethiopia