Semiu Akanmu

(701) 793-8843 | Fargo, ND | ayobami.sm@gmail.com | https://github.com/Semiu

SKILLS

Languages and Scripting | Python3, SQL, Javascript , Java, Linux, Ubuntu, Windows PowerShell

Databases | MySQL, PostgreSQL, Apache Cassandra, MongoDB (with Mongoose and Cloud Atlas), SQLite

Application Frameworks | Django, Flask, PySpark, StreamLite, Express JS, D3.js, NodeJS

ML | MLflow, Scikit-Learn, TensorFlow, Fastai, NLTK, OpenCV, SpaCy, BERT

Cloud and Distributed Data Tech | Apache Spark, AWS Data and ML stack, GitHub, Jenkins

EXPERIENCE

International Consulting Associates

Software Engineer - Data and Machine Learning   **Aug 2022 - Present**

- Leading code quality assurance of a team of 15+ data scientists to achieve production standard code quality.
- Designing, building, testing, deploying and monitoring pipelines for data (extraction, processing and analytics) and ML models (NLP and classification) using AWS ML and data engineering stack.
- Processing, transforming, modeling textual data and identifying Name Entities in texts using SpaCy, NLP toolkit, Pandas and Python to improve (time and data quality) data analytics, feature engineering and machine learning modeling process.

North Dakota State University (Dept of Comp. Sci)

Machine Learning Research Assistant   **Aug 2018 - Present**

- Exploring machine and deep learning models, especially Bi-LSTM, Graph Neural Network and XGBoost, for automated detection of security vulnerabilities in Java program source code.
- Training text classification model using NLP vectorization technique (CountVectorizer), TF-IDF, with Scikit-Learn and Tensorflow for Java CVEs.
- Implemented NLP approach using stop words for text pre-processing, NLTK, Lemmatizer, Tokenizer, and other Python libraries to identify Java CVEs in the MITRE CVE descriptions. Code Repo URL: shorturl.at/xIL57
- Identified the important features for ~70% accuracy prediction of malware susceptibility in Window machines using Scikit-learn libraries in training XGBoost, Random Forest, Logistic Regression and MLP models. Code Repo URL: shorturl.at/jxHUY
- Developed ML (decision tree and Logistic Regression) models using Python and Scikit-learn to predict loan default in a P2P lending community with ~97% accuracy. Code Repo URLs: https://github.com/Semiu/Lender

Eagle Creek Software

Associate Consultant - Data and Machine Learning Engineering   **Aug 2021 - Aug 2022**

- Productivized ML applications using TensorFlow, Sci-kit Learn, Fast API, and Python Django framework to achieve different business use cases, and support clients' demands and revenue.
- Implemented an ETL pipeline using Python, Pandas, AWS-Lambda functions, Glue jobs, Step functions and Snowflake in AWS to improve the data analytics and machine learning process.

NDSU Dickinson Research Extension Center

Software Engineer – Business Intelligence   **Aug 2018 - Dec 2020**

- Developed a business intelligence system using Node.js, Express.js, and SQL in MySQL to achieve actionable reports in a multi-user distributed secure platform supporting $2m revenue beef production.
- Designed a normalized data model in MySQL to remove data redundancy by 80%.
- Developed optimized SQL queries in MySQL to generate business intelligence reports.
- Implemented ETL pipeline using Java Server Pages, Node.js, jQuery scripts/modules, and SQL in MySQL for data input sanitization, automated data cleaning, and bulk data upload.
- Collaborated with team members in code review, merging, and deployment using Git in GitHub and Jenkins to achieve CI/CD for product delivery improvement.
- Performed requirement gathering sessions and user-acceptance testing with the ND Beef producers.

Augustine University

Database Instructor   **Sept 2017 - July 2018**

- Taught theoretical and hands-on practical courses on data warehouse (using MS SQL Server), and database design and modeling (SQL and MySQL).

Durban University of Technology

Data Scientist - Visiting Research Associate   **Sept 2016 - Aug 2017**

- Built an automated and distributed ETL for machine learning pipeline using Python in PySpark to improve the pre-processing, cleansing, and storage of dam operational data to 95% efficiency.
- Implemented Multilayer Perceptron model stacked with XGBoost with Scikit-learn and Python to achieve ~90% accuracy of flash flood prediction.

Bacamax Sdn Bhd (Defunct)

Application Developer - Data Visualization   **October 2014 - July 2016**

- Developed data visualization system using NodeJS and D3.js, Chart.js, SQL, MySQL to (i) interact with IoT APIs, (ii) visualize real-time vehicular movement with 100% accuracy, and (iii) calculate drivers' safety compliance and job performance rate.
- Collaborated with transport and logistic professionals to achieve a user-centered actionable dashboard supporting RM120 million in business decision-making.

EDUCATION

North Dakota State University

Ph.D. - Doctor of Philosophy in Software Engineering, ABD.   **August 2023**

M.S. – Masters of Science in Software Engineering   **June 2022**

Universiti Utara Malaysia

M.S - Masters of Science in Information Technology   **February 2013**

Olabisi Onabanjo University, Nigeria

B.S. - Bachelor of Science in Computer Science   **June 2008**