CHUKWUMA HILARY AKPU

+447769380756 | holaryc@gmail.com
github.com/hilariie | linkedin.com/in/hilary-ml-engineer

PROFESSIONAL SUMMARY

An experienced data scientist with 4 years of work experience applying data science to business applications. Particularly experienced in developing and deploying high-performing machine learning and deep learning models to solve business problems. I am rooted in the pursuit of impacting lives and businesses through my experience and skills by developing solutions founded on data.

TECHNICAL SKILLS

Machine Learning: Logistic/linear regression, random forest trees, K-means, gradient boosted trees.
Deep Learning: Natural language processing, Convolutional Neural Network, BERT (LLMs), VGG16, Transformers, GPT, Whisper, SAM, YOLO, DeepSORT, Image segmentation.
Libraries: NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, ggplot, TensorFlow, Beautiful Soup, Scrapy.
Programming: Python, R, Object-oriented programming, SQL, Version Control (Git), CI/CD, unit testing, APIs.
Tools and Technologies: Apache Airflow, Microsoft Azure (Databricks, Data Factory), KNIME, Microsoft Excel, Docker, Anaconda, PyCharm, Apache Spark, Hadoop, PowerBI, Octoparse.
Database: Snowflake, PostgreSQL, CouchDB (NoSQL).

WORK EXPERIENCE

Research Volunteer • University of Reading **Jun. 2023 – Present**

- In collaboration with UK scientists, I developed a multi-object tracking and analysis system using YOLO and DeepSORT that provides valuable soil information. *(github link)*
- Working on publishing a paper on the automated tracking and analysis system I developed.
- Researched, proposed, and implemented modifications to state-of-the-art algorithms.
- Developed an audio transcriber using OpenAI's Whisper and GPT models to help people in rural areas get better access to vital information. *(github link)*
- Developed algorithms that distinguished between male/female animals and detected mating occurrences in a video using image segmentation coordinates and the Euclidean Distance metric.
- Utilized adaptive thresholding and the Segment Anything Model for precise image segmentation.
- Annotated 500+ images for image detection model training.

Data Scientist • Reliance HMO **Jun. 2022 – Oct. 2022**

- Trained BERT and sequential (CNN for text) models to classify diagnoses with an accuracy above 90%. This improved the labelling rate by 10% and provided a foundation for improving the pricing model to reduce loss. This also brought to light discrepancies in the data that were otherwise hidden.
- Led the pre-processing of a large text dataset using techniques such as lemmatization, stemming, and regex pattern matching.
- Developed a labelling algorithm to accurately label a healthcare dataset, resulting in a 70% labelling rate.
- Spearheaded the development of a user-friendly application to access the pricing model prototype using CSS, HTML, and Python.
- Executed data engineering tasks on Snowflake using SQL queries.
- Leveraged PowerBI to analyze and visualize insights, providing valuable information for executive decision-making.

Data Scientist • Ehealth4Everyone **Dec. 2020 – May 2022**

- Conducted advanced customer segmentation (B2B) using firmographic segmentation and tiering, boosting customer satisfaction.
- Successfully migrated a database from CouchDB to PostgreSQL for an application by re-architecting the codebase and executing data engineering tasks on unstructured data.
- Developed a notification system for a mobile app, increasing vaccination rates by 25%.
- Automated staff productivity tracking process, reducing manual input by 100%.
- Conducted data collection, pre-processing, and automation using Apache Airflow on Docker.

Data Analyst • Softcom Limited **Apr. 2019 – Oct. 2020**

- Conducted rigorous data analysis and created reports using Power BI, Excel, and other data visualization tools to support business decisions.

EDUCATION

MSc Data Science – University of Reading – Reading, England. **Sep. 2022 – Sep. 2023**

- Graduated with a Distinction.
- Big Data: Developed and deployed machine learning models on Azure and PySpark.
- Cloud Computing: Developed a concurrent MapReduce algorithm without relying on Hadoop *(github link)*.
- Mathematics and Statistics: Deepened my understanding of the foundational and statistical aspects of data science.
- Data ethics and security

BSc Computer Science – Mountain Top University – Ogun, Nigeria. **Oct. 2016 – Dec. 2020**

- GPA 4.48/5.0.
- Completed a dissertation on using machine learning models for diagnosing heart disease with an accuracy above 90%.

Computer vision Nanodegree – Udacity. **Sep. 2020**

- Developed a CNN model to detect key points on the face.
- Developed an encoder-decoder model that generates captions for images.

PERSONAL PROJECTS

Multi-object tracking **May 2023**

- Developed a program using image segmentation and OpenCV's tracking algorithms, to track multiple objects in a video. *(github link)*

Python emailing package **Mar. 2022**

- Designed and developed a comprehensive Python package that simplifies the process of sending emails, achieving over 8,000 downloads in a year. *(github link)*

Multi-output Neural Network **May 2023**

- Developed a system to simplify the implementation of the one-vs-all approach on multiclass image problems. I made modifications to classes and functions from the TensorFlow library indicative of an improved understanding of the library and neural network in general. *(github link)*