HILARY AKPU

+447769380756 | holaryc@gmail.com | github | portfolio

PROFESSIONAL SUMMARY

A data scientist with 4 years of experience and an MSc degree in data science. My experience working as a Data Analyst, a Software Engineer, and a Machine Learning Engineer has equipped me with the skills and knowledge to tackle a wide range of data science problems others may struggle to tackle. I am eager and excited to put my skills to use and solve difficult problems for your organisation.

SKILLS

Python | Computer Vision | Neural Networks | NLP | SQL | APIs | TensorFlow | PyTorch | Azure | Airflow | Software Engineering | Research | Data Analysis and Visualisation | Storytelling.

WORK EXPERIENCE

University of Reading **Jun. 2023 – Mar. 2024**

Computer Vision Researcher

- Collaborated with a team of distinguished professors to generate and submit a paper to the ICPR conference as the first author.
- Researched, developed, and implemented an automated object tracking and analysis system using state-of-the-art computer vision algorithms.
- Utilized adaptive thresholding and SAM for precise image segmentation to aid object tracking.
- Utilised the Euclidean Distance metric to develop custom algorithms that provided new solutions to difficult problems.
- Conducted statistical analysis to evaluate model performance.
- Provided accurate time estimates for project deliverables, delivering code on schedule and advocating for quality coding practices.

Tools Used: *Python, PyTorch, TensorFlow, CNN, Image processing and segmentation, YOLO, DeepSORT, Statistics, Mathematics (MOTA, Euclidean Distance), OpenCV, Numpy, Git.*

Reliance Health **Jun. 2022 – Oct. 2022**

Data Scientist

- Collaborated in the development of a pricing model that reduced the Medical Loss Ratio.
- Utilised state-of-the-art LLMs like BERT to classify large datasets with an accuracy above 80%.
- Developed a text-labelling algorithm that aided the development of machine learning models.
- Led the pre-processing of a large text dataset using techniques such as lemmatization, stemming, and regex pattern matching.
- Spearheaded the development of a user-friendly application to access a pricing model prototype using CSS, HTML, and Python.
- Presented findings and recommendations to senior leadership and stakeholders, demonstrating the value of data-driven solutions.

Tools Used: *Python, TensorFlow, Transformers, LLMs, BERT, Git, CSS, HTML, Django/Flask, APIs, Github, Snowflake.*

Ehealth4Everyone **Dec. 2020 – May. 2022**

Python Data Scientist

- Developed and managed automated data pipelines using Airflow DAGS.
- Saved the lives of children by developing a notification system for a mobile application which increased vaccination rates.
- Migrated the database of a mobile application from NoSQL (CouchDB) to PostgreSQL.
- Crafted algorithms to monitor the status of data pipelines and notify stakeholders accordingly.
- Developed scraping scripts to scrape relevant datasets from free-to-scrape websites.
- Automated staff productivity tracking process, reducing manual input by 100%.

Tools Used: *Python, Pandas, Beautiful Soup, Scrapy, SQL, PostgreSQL, Airflow, Docker, Git, Gitlab, API, Unit testing, Code reviews.*

Softcom **Apr. 2019 – Oct. 2019**
Data Analyst Intern

- Conducted geo-spatial analysing highlighting the concentration of outlet stores while showing potential areas for expansion.
- Analysed results from questionnaires to improve the company's understanding of their customers while also suggesting actions to be taken for improvements.
- Performed regular data cleaning, analysis and visualisation tasks using Excel and Python.

Tools Used: *Python, Pandas, Excel, Numpy, Pandas, Matplotlib, seaborn, Jira.*

EDUCATION

MSc Data Science – University of Reading. **Sep. 2023**
BSc Computer Science – Mountain Top University. **Dec. 2020**
Computer Vision Nanodegree – Udacity. **Sep. 2020**

PERSONAL PROJECTS

Audio Transcriber (Gen AI)

- Designed and implemented a video translator system utilizing Gen AI algorithms for transcription and translation, contributing to the advancement of NLP.

Open-Source Emailing Library

- Developed a Python emailing library with 11,000 downloads in its second year of release. (*code*)

Computer Vision Library

- Developed a system to simplify the implementation of the one-vs-all approach in image classification.
- Developed custom data augmentation and batching algorithms to suit the OvA approach.

Breast Cancer Research

- Conducted research on developing a breast cancer diagnosis system with a recall of 99%.