Samuel Bodza

Data Wrangler

Contact

📍 BS6 6SA, UK

✉ Sambodza@hotmail.co.uk

📱 +(44)7794708855

in [Samuel Bodza](#)

🐙 [Bonsa11](#)

Technical Skills

Languages

- Python
- Bash
- SQL
- R

Python Packages

- Pandas
- Numpy
- OpenCV
- Sci-Kit
- TensorFLow
- Keras
- Flask
- Pytest
- Streamlit

Technologies

- Docker
- Docker Compose
- PostgreSQL
- SQLServer
- Git
- PowerBI
- Linux
- Ubuntu
- CentOS

Education

Msc, Data Science - **Distinction**
09/19 - 09/20
Liverpool John Moores University

Bsc, Physics - 2:2
09/16 - 09/19
University of Southampton

References available upon request.

Summary

I am a UK based Data Wrangler. I have extensive experience designing, writing, and implementing end-to-end data pipelines that contain custom deep learning algorithms. I enjoy the challenge of writing production ready code to implement tools and services that help others complete their work

Experience

Data Wrangler | University Hospital Birmingham | 02/2021 - Present

HIGHLIGHTED PROJECT: End-to-End Design and Development of Data Pipeline

- 15Tb of images to; scrape and strip metadata, remove identifiable images and change format to DICOM

- Designed and implemented data pipeline that was required to be built around software from external company to convert images into dicoms.

- Trained, tested, and incorporated two CNNs in pipeline. 99.8% accuracy for removing 'identifiable' images and 98.8% accuracy for feature extraction to fill NULL values in structured data in PostgreSQL.

- Each stage of pipeline was an isolated docker container, if a container would run was controlled by logic in a PostgreSQL table

Building Custom Tools

- Built automated tools for Data Developers to be able to easily access resources that required manual intervention beforehand.

Reverse engineering

- Reverse engineered a proprietary file type and scraped files for IDs that linked back to an existing database which allows us to link across different data sources which were inaccessible before.

Infrastructure Management

- Built and manage Postgres databases and Ubuntu application servers. Independently incorporated uptime monitoring of on-prem servers, with automated notifications that were pushed through to a Teams channel with other members of the infrastructure team.

Team Practices

- Pushed for sharing code between teams in GitHub to reduce repeated work, pushed to implement coding standards such as unit testing, standardised logging, and error handling for all production code.

Freelance Work | OCR Data Extraction for Academic Paper

- Tasked with extracting data from scanned PDFs. PDF files required thorough cleaning and OCR required complex logic. Paper "Is point wise analysis of the Humphrey visual field feasible as a primary outcome in idiopathic intracranial hypertension?" currently in the process of being published in TVST.

Personal Projects | Controlling Drone using Webcam

- Using OpenCV, a TensorFlow CNN was combined with a Decision Tree and were trained to identify the positions of my hands, this was then used instead of the controller to pilot the drone.