Habeeb Shopeju

Software Engineer, Machine Learning

Website | Github | LinkedIn | shopejuh@gmail.com | London, United Kingdom

Software Engineer with Flair for Machine Learning Application and Research

Software Engineer with 4 plus years of experience in building machine learning models, data processing pipelines and enabling research that make machine learning valuable to people and businesses. Excited about building information retrieval and machine learning solutions.

Relevant Tools

- Programming Languages: Python | Java | C++
- Machine Learning: Pytorch | Sci-kit Learn | Pandas | Numpy | ONNX | Backtrader | Huggingface Libraries | Jax
- Backend: Django | FastAPI
- Infrastructure: Docker | Kubernetes |Apache Beam | AWS SageMaker | Azure Machine Learning
- Databases: Postgres | Neo4j | Redis | Elasticsearch | Milvus

Work Experience

Thomson Reuters Labs, London | Research Engineer, Machine Learning | 2022 – Present

- **Main tools: Pandas, Huggingface, Elasticsearch, Milvus, AWS, Azure, OpenAI, TR Labs tools.**
- Building cloud tools: Making contributions to the internal tools used by the Research Scientists when using AWS Sagemaker and Azure Machine Learning, the goal is to ease and speed up the experimentation-to-production workflow.
- Building AI Legal Clause Drafting Assistant: Setup the research environment used by Research Scientists for testing different Retrieval Augmented Generation (RAG) ideas. Making algorithms production-ready and integrating them into the product's backend service.
- Refactored BERT models from an acquisition: Refactored and helped debug existing BERT model packages after an acquisition so they are compatible with the existing Machine Learning setup at the Labs.

Primer.ai, London | Software Engineer, Machine Learning | 2021 – 2022

- **Main tools: Pandas, FastAPI, Huggingface, PostgreSQL, Neo4j, Apache Beam, SPARQL, Wikidata, AWS, GCP**
- Built data pipeline for knowledge base: Built ETL process for self-updating Machine Learning Platform's knowledge base that augments documents for company's named entity extraction and resolution tasks.
- Helped build an entity search engine: Contributed to building the data ingest pipeline for an entity search engine that is powered by an ensemble of Primer's in-house Machine Learning models.

7Q1, Stuttgart (Remote) | Software Engineer, Backend & NLP | 2020 - 2021

- **Main tools: Vanilla Javascript, Django Rest Framework, NLTK, PostgreSQL, AWS.**
- Built data collection tool: Built a chrome extension for extracting company details from web pages; all data scientists and data entry personnel at the company used this tool to get and modify company data before storing in the database.
- Core engineer on APIs: Played an integral role in designing the database architecture and building the APIs for a supplier search engine that reduces procurement process costs for customers by 75%.

Wikimedia Foundation, San Francisco (Remote) | Machine Learning Research Intern | 2020

- **Main tools: Spacy, Gensim, NLTK, Regex, Wikimedia tools.**

- Improved text tokenization: Increased efficiency of regular expressions and rules used for text tokenization in ML pipeline, reducing time taken by 2x without regression in tokenization quality. Added detection of foul words and idioms in English Wikipedia to improve feature extraction for topic modelling.
- Gained and presented insights about Wikimedia topic models: Analyzed and compared the results of existing Wikimedia language-agnostic topic models, doing this research on Arabic, Czech, English and Vietnamese articles.

TrueBeauty, Warsaw (Remote) | Software Engineer, Computer Vision | 2020

- **Main tools: Numpy, Scipy, Detectron2, Tensorflow Object Detection, Minio.**
- Built an image tagging pipeline: This pipeline was used to tag various facial features, augment the features with metadata then save them in the object store; the pipeline was used for feature retrieval too.
- Worked on the core logic of cosmetic surgery software: Researched various beauty criteria and worked on logic for implementing such cosmetic changes on facial images as well as generating beauty scores of faces based on those criteria.

Xerde, Lagos | Software Engineer Intern, Backend | 2019 – 2020

- **Main tools: Django Rest Framework, PostgreSQL, Redis, Celery, Nginx.**
- Built the backend of a Fintech service: I built and wrote tests for multiple components of a platform that helps people save and seek contributions towards their goals.

Side Projects

Trading Stocks using Machine Learning: Combining machine learning capabilities with price reversion concepts to trade stocks. Building it with tools like Backtrader, Scikit-Learn, OpenBB. I also reverse-engineered Trading212 APIs to programmatically send trading requests.

Reddington: A complaints discovery prototype I built to extract complaints made on Reddit, so indie hackers searching for problems can easily find them. Building it with tools like Streamlit, VADER Sentiment, FAISS, OpenAI, Minio, Yet Another Keyword Extractor (YAKE).

Education

2021 | Bachelor of Engineering, Electrical and Electronics Engineering | Federal University of Agriculture, Abeokuta