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

Engineer with 4 plus years of experience in building machine learning models, services and data pipelines to make machine learning valuable to people and businesses via collaborations with scientists and domain experts. Flourishes in environments that encourage asking questions. Excited about building information retrieval and machine learning solutions.

Relevant Work Experience

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

- Significantly used: Huggingface Libraries, XGBoost, Pandas, Opensearch, Milvus, Apache Spark, Docker, FastAPI, OpenAI, AWS, Azure, Internal Tools.
- Building cloud tools: Worked on internal tools to aid experimentation with cloud services. Integrating features like a Streamlit prototyping functionality and profile management across multiple projects and experiments. **Constantly working on feedback to improve this tool used by over 100 scientists and engineers across the Labs.**
- Building a contract drafting service: Worked with scientists to run experiments on Retrieval Augmented Generation ideas for the functionality. Worked on the data indexing pipeline, vector database infrastructure and backend service that **helps lawyers generate contracts in the company's flagship Generative AI product.**
- Building an information extraction service: Working on experiments for accurately extracting semantically meaningful sections from standard documents and making the solution into an internal service. **This functionality is fundamental to at least 3 teams building comparison services across the labs.**

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

- Significantly used: PyTorch, Tensorflow, Huggingface Libraries, Pandas, NetworkX, SPARQL, Wikidata, PostgreSQL, Neo4j, Apache Beam, FastAPI, AWS, GCP.
- Built data pipeline for a knowledge base: Helped set up the knowledge base and build the pipeline for updating the knowledge base using Wikidata and various news sources. Worked on the service used for getting entity metadata, **serving other teams working on named entity extraction and resolution tasks**.
- Built an investment returns analytics tool: Wrote parsing logic and processing pipeline to extract financial data from PDFs containing investment returns data and store in the database. **Worked on the API that serves this structured data to the user interface as a custom solution for one of the company's most trusted clients**.

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

- Significantly used: NLTK, Spacy, PostgreSQL, Vanilla Javascript, Django, AWS.
- Built a procurement search engine: Helped build a product for finding suppliers by providing names of goods and equipments. Worked on the database architecture design, built the crawler manager and search logic, **reducing procurement process costs for the company's biggest German manufacturing clients**.
- Built a data collection tool: Built a chrome extension for extracting supplier details from web pages and save in a database. **Learnt Javascript to build this tool that brought substantial improvements to the workflows of all data entry personnel on the product**.

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

- Main tools: Spacy, Gensim, NLTK, Regex, Wikimedia tools and custom infrastructure.
- Worked on article integrity scoring: Added detection of foul words and idioms in English Wikipedia to improve feature extraction for topic modelling. Increased efficiency of regular expressions and rules used for text tokenization in ML pipeline, **reducing time taken by 2x without regression in tokenization quality**.
- Analyzed the quality of topic models: Collaborated with a research scientist to analyze the quality of the company's three topic models, doing this research on Arabic, Czech, English and Vietnamese articles. **Communicated the insights found to key stakeholders, influencing the decision of the model to invest more resources into**.

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

- Main tools: Numpy, Scipy, Detectron2, Tensorflow Object Detection, Minio, Streamlit
- Built an image pipeline, cosmetic surgery prototype: The pipeline to helped to tag, augment and retrieve facial features. Contributed a prototype for applying cosmetic changes on faces using pipeline features.

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

- Main tools: Django Rest Framework, PostgreSQL, Redis, Celery, Nginx.
- Worked on a savings product: I built and wrote tests for multiple components (email notification, referrals, search, payments) of this product that helps people save and seek contributions towards their goals.

Recent Side Projects & Activities

- Learning Rust, Jax, ONNX.
- PyTensor: Open-source contributions, e.g. Implementing PyTensor Ops using Pytorch.
- Reddington: Building a tool to convert complaints on Reddit into user statements, useful for product managers.
- Trading Stocks using Machine Learning: Combining machine learning capabilities with price reversion concepts to trade stocks.
- tradingTOT: Reverse-engineered Trading212 APIs to programmatically send trading requests.

Education

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