

Atharva Niranjn Joshi

atharvanj@gmail.com | +44 – 07587039921 | 45C Mawney Road, Romford, United Kingdom RM77HL | [linkedin.com/in/atharvanj](https://www.linkedin.com/in/atharvanj) | atharvanjoshi.github.io/

SUMMARY

I am a Forensic Data Analyst working at KPMG UK specifically in the Unstructured Analytics department. I use Data and Machine Learning for Trader Surveillance. I have completed my MSc. in Data Science from QMUL. I also worked as an AI Intern at CHRE using Computer Vision to improve Oral Cancer diagnosis. I wish to bring a positive change to the world using the power of data. I am originally from India and I love to read and play cricket and badminton.

WORK EXPERIENCE

Graduate Forensic Data Analyst, KPMG, UK

Oct 2022 – Present

- Developed NLP TF – IDF Machine Learning model using SGD classifier for text and audio communications to detect suspicious activities in Trader Surveillance. Experimented with other models using TensorFlow Library.
- Deployed models on Azure Cloud Platform server using Batch Inference Pipeline
- Developed a tool using GPT Large Language Model (LLM) provided by Microsoft for reviewers for easy detection of fraud in communications using few shot inferences and various techniques to improve like instruction fine tuning and parameter efficient fine tuning (LoRA and Soft prompts). Currently exploring Fine Tuning with Reinforcement Learning and Program Aided Language Models to query data.
- Worked with OpenAI Whisper models to improve upon existing Text 2 Speech Transcription Service accuracy.
- Used Python extensively for data engineering, exploratory data analysis, data cleaning, transcription, translation, data ingestion and storage, programming and debugging, process automation and for an end – to – end data pipeline. Working to create a CI/CD pipeline and host it on Azure.
- Developed a tool using Python, Sequence Matcher and NLP to find and remove duplicate communications.
- Used Microsoft Excel for analysis of data as well as creating appropriate dashboards and reports of model performance and monitoring data.
- Used SQL for regular tasks such as Ingesting and Sampling of Data into Relativity platform.

AI Consulting Intern, Centre for Health Research and Education, UK

Jun 2021 – Aug 2022

- Researched and applied various Image object detection and classification models like YOLO v4 and R -CNN using Python and Deep Learning frameworks like PyTorch on existing as well as newly created data.
- Coordinated with various medical institutions for obtaining data, managing and storing on Amazon Web Services (AWS) S3 buckets cloud platform for oral cancer research connecting that with a standalone version of LabelImg for annotation.
- Coordinated with medical experts to understand the domain and in creating appropriate datasets.

Systems Engineer, Infosys Limited, India

Sep 2018 – Sep 2020

- Developed applications in Android and iOS using Xamarin.Forms framework based on C# (Object Oriented Programming) and XAML using DevOps on Microsoft Visual Studio.
- Used Azure Active Directory for authentication and Azure APIs for data retrieval and storage.
- Played an important role in the development cycle by implementing complex features, coordinating with my teammates, tracking the progress of each use case development and bug fixes through the Azure DevOps Portal.
- Developed Pilot projects and also presented Proof of Concepts, the working of apps through emulator and walkthrough of codes to the clients through Microsoft Teams.
- Nominated and awarded with a Certificate of Appreciation for my performance in the company.

EDUCATION

Master of Science (MSc.), Big Data Science with Industrial Experience
Queen Mary University of London

Sep 2020 – Sep 2022
Percentage: 86.31 (Distinction)

Post Graduate Diploma, Data Science
IMS ProSchool, India

Apr 2019 – Feb 2021
Grade: A

Bachelor of Engineering (B.E.), Information Technology
Savitribai Phule Pune University

Jul 2014 – Jul 2018
Percentage: 63.61 (2:1)

12th (Equivalent to A – Level), Computer Science
Maharashtra State Board, India

Feb 2013 – Feb 2014
Percentage: 71.69

10th (Equivalent to GCSE)
Maharashtra State Board, India

Mar 2011 – Mar 2012
Percentage: 84.55

CERTIFICATIONS

- Infosys Certified Python Programmer and Python Associate
- Natural Language Processing with Classification and Vector Spaces – Coursera
- Generative AI with Large Language Models – Coursera
- Natural Language Processing Specialisation – Coursera (In Progress)

SKILLS

Data Science: Python, SQL, Excel, Relativity

Languages: English (Fluent, C1(IELTS)), Marathi (Fluent, Mother Tongue), Hindi (Intermediate)

PUBLICATIONS

E-Health Monitoring and Detection using Wireless Sensor Networks 2018 - *International Engineering Research Journal*