Shobhit Maheshwari

shobhit.workds@gmail.com | +44-7826679948, +91-9599236447 | Portfolio | Linkedin | Github

SUMMARY

Data Scientist with 4.5+ years of experience specializing in Computer Vision and NLP, underpinned by a strong foundation in Mathematics. Skilled in Python and SQL, with extensive hands-on use of leading ML frameworks such as PyTorch and TensorFlow. Currently working on optimising QA frameworks like RAG and MapReduce for processing long financial documents. Seeking opportunities that encompass the full model lifecycle, from ideation to delivering impactful solutions.

SKILLS

Python, SQL, Pandas, Matplotlib, Numpy, OpenCV, Scikit, Keras, TensorFlow, PyTorch, PySpark, LangChain, NLTK, Spacy, Selenium, BS4, Flask, FastAPI, Git, Docker, Kubernetes, Nginx, AWS, Airflow, Jenkins, Metabase, A/B testing, Statistics

EXPERIENCE

Research Assistant, *University of Edinburgh | Edinburgh, UK*   *Aug 2024 – Present*
- o   Benchmarked GPT against 4-bit quantised Llama and Mistral on FinQA, demonstrating a ~12% gain in accuracy.
- o   Achieved a 16% improvement over RAG using a ranked MapReduce pipeline with similar processing time to RAG.
- o   Scraped financial data sources to compile 70GB from annual reports, earnings transcripts, etc.; developing a low-latency framework to query large financial reports for all companies on a unified platform.

Senior Data Scientist, *Roadzen | Delhi, India*   *Jan 2022 – Aug 2023*
- o   Streamlined ML workflow from data ingestion to model serving using FastAPI, Airflow, and Jenkins, resulting in continuous training and efficient deployment of the in-house ML models.
- o   Enhanced the accuracy of the insurance policy QA bot using GPT, achieving 95% correct answer retrieval.
- o   Developed a heuristic inspired by the concept of momentum to extract keyframes for car profiles from a video.
- o   Achieved over 84% accuracy in classifying car colour, make, and model for expedited claims validation.

Data Scientist*, Roadzen | Delhi, India*   *Jun 2020 – Dec 2021*
- o   Built a mask RCNN model for instance segmentation of damage, parts, and profile on car images using PyTorch, reducing claims processing time from 40 mins to under 2 mins.
- o   Optimized the mAP score metric to account for subjectivity, achieving a score of 74 in damage segmentation.
- o   Reduced Damage Recognition API turnaround time by 30% using torchserve and FastAPI.
- o   Earned recognition for AI models at the Asia Motor Insurance Summit and Financial Express Future Tech awards.

Data Scientist, *Spoonshot | Bangalore, India*   *Jan 2019 – Jun 2020*
- o   Designed a weighted DeepWalk model with ingredients as nodes and edge traversal probability based on flavour pairing theory to generate novel flavour pairings.
- o   Managed 100 million ingredient combinations by building a PySpark pipeline for efficient edge weight generation.
- o   Implemented Fast-RCNN to extract nutrition panel with a mAP of 85 from product images.
- o   Utilized Azure OCR to extract nutrition components and their corresponding values with 93% accuracy.

EDUCATION

- **University of Edinburgh** | *Edinburgh, UK*   *Sep 2023 – Nov 2024*
  M. Sc. In Data Science   *Grades - Distinction*
- **University of Delhi** | *Delhi, India*   *Aug 2015 – Jun 2019*
  B. Tech. in Information Technology and Mathematics   *Grades - 89.5/100*

PROJECTS

Neural Machine Translation (NMT)
- o   Implemented self-attention from scratch for German to English, generating BLEU-4 of 13.5 with 10k samples.
- o   Enhanced decoding with a lexical module, gaining a 2.6-point BLEU improvement over greedy decoding.

X-ray Image denoising using U-Net
- o   Generated noisy NIH dataset images by applying noise masks multiplied with a 256 x 256 Gaussian mask.
- o   Developed a U-Net x-ray denoising model with short and long range skip connections, achieving a PSNR of 27.