Shobhit Maheshwari

Data Scientist, Machine Learning Engineer

96 Holyrood Rd, Edinburgh EH8 8FH | shobhit.workds@gmail.com | +44-7826679948 | Linkedin | Github

Data Scientist with a strong foundation in mathematics and 4 years of work experience in Computer Vision and Natural Language Processing (NLP). Adept in Python and highly skilled with advanced Machine Learning libraries like TensorFlow, PyTorch, and LangChain. Demonstrated success in developing and deploying high-impact solutions in production environments.

EDUCATION

University of Edinburgh *(Ranked 27th in the 2025 QS rankings)* Sept 2023 - Present
M. Sc. in Data Science (Distinction grades in 90/120 credits)

- **Modules:** Natural Language Processing, Machine Learning Practical, Extreme Computing, Image and Vision Computing, Natural Language Understanding, Pattern Recognition in Financial Data
- **Dissertation Project (QA framework for long financial PDFs):**
  - Historical annual reports scraped for FTSE all-share accounting for 70GB data.
  - Implemented Map Reduce with ranking to perform Question Answering for extracting business insights.
  - 16% improvements in results with equivalent processing time compared to RAG.

University of Delhi *(Ranked 11th in Indian institutional rankings)* Aug 2015 - Jul 2019
B.Tech in Information Technology & Mathematical Innovation (89.5/100)

WORK EXPERIENCE

Senior Data Scientist, Roadzen | Delhi, India Jan 2022 - Aug 2023

- Streamlined ML workflow from data ingestion to model serving using FastAPI, Airflow, and Jenkins in collaboration with Engineering team.
- Developed insurance policy QA bot with GPT to significantly reduce document search times.
- Extracted keyframes for profiles of a car from video using a heuristic inspired by the concept of momentum.
- Achieved over 90% accuracy in classifying car color, make, and model for expedited claims validation.

Data Scientist, Roadzen | Delhi, India Jun 2020 - Dec 2021

- Developed and trained a mask RCNN for instance segmentation of damage, parts, and profile on car images using PyTorch framework Detectron2, reducing claims processing time from 40 mins to 2 mins.
- Achieved custom mAP score (to account for damage subjectivity) of 74 for damage segmentation.
- Reduced Damage Recognition API turnaround time by 30% using torchserve and FastAPI.
- Received accolades at the Asia Motor Insurance Summit and Financial Express Future Tech Awards.

Data Scientist, Spoonshot | Bangalore, India Jul 2019 - Jun 2020

- Developed ingredient embeddings with weighted DeepWalk using ingredients as nodes and edge traversal probability based on flavour pairing theory and ingredient co-occurrence.
- Handled 100 million ingredient combinations using a PySpark pipeline for edge weights generation.
- Leveraged ingredient embeddings for creating unique ingredient pairs recommended to clients including P&G and Coca-Cola.
- Implemented Fast-RCNN to extract nutrition panel and used Azure OCR to get standardized nutrition information from various Consumer Packaged Goods products.

SKILLS

Programming Languages	Python, SQL, HTML, CSS, C/C++
ML, NLP, Computer Vision	Regression, Dimensionality reduction, Random Forest, XGBoost, Adaboost, Clustering, Embeddings, RNN, LSTM, Encoder-decoder models, Transformers, LLM, BERT, Retrieval Augmented Generation (RAG), Vector DB, Feature extraction (HOG, SIFT, LBP, Harris, Canny), CNN, GAN
Data and ML Libraries	Selenium, Scrapy, BeautifulSoup, NLTK, Spacy, Pandas, Seaborn, Matplotlib, Scikit-Learn, Keras, TensorFlow, PyTorch, PySpark, LangChain
Deployment and DevOps	Git, FastAPI, Flask, Docker, Airflow, Jenkins, AWS, Kubernetes, Nginx
Soft Skills	Problem-solving, Teamwork, Communication, Analytical and Creative thinking

PROJECTS

X-ray Image denoising using U-Net

- An x-ray image denoising model built using U-Net architecture using long and short range skip connections.
- Reached PSNR of 27 on the noise induced using Gaussian mask with blur, speckle, etc. on NIH dataset.

Neural Machine Translation (NMT)

- implemented Multi-headed attention from scratch to translate German to English, reaching BLEU-4 of 13.5.
- Comparative study done on greedy decoding, with a lexical decoding module using attention plots.

Analysis of Classical and Deep Learning approaches to image perturbations

- Used HOG and SIFT representations to take a SVC with accuracy of 10% to 61% on Sports ball dataset.
- Studied the effects of varying intensities of perturbations on hand-crafted and features extracted by ResNet.

Image Captioning using Flickr8K

- Trained an end-to-end model on Flickr8k data with encoder architecture of VGG16 and an LSTM decoder.
- Decapitated the final fully connected layers from encoder to generate captions reaching BLEU-1 of 58.

RESEARCH AND PUBLICATIONS

- Mahima Kaushik*, Shobhit Maheshwari and Rddhima Raghunand, “Exploring Promises of siRNA in Cancer Therapeutics”, Current Cancer Therapy Reviews (2019) 15: 1. https://doi.org/10.2174/1573394715666190207130128
- Maheshwari, Shobhit, and Rddhima Raghunand. “Multi-Character Recognition using EMNIST.” JIMS8I-International Journal of Information Communication and Computing Technology 6.1 (2018): 325-331

EXTRA CURRICULARS

- Lead organiser of DataJam, a Data Science event hosted by the college, collaborating with a team of five.
- Social Media Manager of Autonomi, an autonomous student run robotics society at college.
- Represented the school table tennis team and played at the State level tournament for Rajasthan.