

# Shobhit Maheshwari

## Data Scientist, Machine Learning Engineer

shobhit.workds@gmail.com | +91-9599236447, +44-7826679948 | [Portfolio](#) | [Linkedin](#) | [Github](#)

*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.*

### WORK EXPERIENCE

**Research Assistant**, University of Edinburgh | Edinburgh, UK

Aug 2024 - Present

- Historical annual reports scraped for FTSE all-share accounting for 70GB data.
- Contrasted GPT with 4-bit quantised Llama and Mistral on QA tasks; GPT outperformed both by ~12%.
- Achieved a 16% improvement over RAG by implementing a MapReduce with Ranking pipeline, generating answers with the same processing time as RAG.
- Currently working towards refining the framework and deploying it for easy QA on long PDF documents.

**Senior Data Scientist**, Roadzen | Delhi, India

Jan 2022 - Aug 2023

- 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.
- Improved efficiency of the insurance policy QA bot with GPT to retrieve correct answers with 95% accuracy.
- Extracted keyframes for profiles of a car from video using a heuristic inspired by the concept of momentum.
- Achieved over 84% accuracy in classifying car colour, make, and model for expedited claims validation.

**Data Scientist**, Roadzen | Delhi, India

Jun 2020 - Dec 2021

- 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.
- Optimised the mAP score metric to account for ground truth subjectivity reaching a score of 74.
- Reduced Damage Recognition API turnaround time by 30% using Torchserve and FastAPI.
- The AI models were awarded at the Asia Motor Insurance Summit and Financial Express Future Tech.

**Data Scientist**, Spoonshot | Bangalore, India

Jan 2019 - Jun 2020

- Designed a weighted DeepWalk model with ingredients as nodes and edge traversal probability based on flavour pairing theory and ingredient co-occurrence to generate novel flavour pairings.
- Handled 100 million ingredient combinations using a PySpark pipeline for edge weight generation.
- Implemented Fast-RCNN to extract nutrition panel with a mAP of 85 from product images.
- Used Azure OCR to extract nutrition components and its corresponding values with an accuracy of 93%.

### EDUCATION

**University of Edinburgh** (Ranked 27<sup>th</sup> in the 2025 QS rankings)

Sept 2023 - Nov 2024

**M. Sc. in Data Science** (Distinction)

**University of Delhi** (Ranked 11<sup>th</sup> in Indian institutional rankings)

Aug 2015 - Jul 2019

**B. Tech. in Information Technology & Mathematical Innovation** (89.5/100)

## SKILLS

---

<b>Programming Languages</b>	Python, SQL, HTML, CSS
<b>ML, NLP, Computer Vision</b>	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
<b>Data and ML Libraries</b>	Selenium, Scrapy, BeautifulSoup, NLTK, Spacy, Pandas, Seaborn, Matplotlib, Scikit-Learn, Keras, TensorFlow, PyTorch, PySpark, LangChain
<b>Deployment and DevOps</b>	Git, FastAPI, Flask, Docker, Airflow, Jenkins, AWS, Kubernetes, Nginx
<b>Soft Skills</b>	Teamwork, Communication, Analytical and Creative thinking

## PROJECTS

---

### X-ray Image denoising using U-Net

- Generated the noisy NIH dataset images by multiplying noise masks with a Gaussian mask of 256\*256.
- Built an x-ray denoising model using U-Net with long and short range skip connections reaching PSNR of 27.

### Neural Machine Translation (NMT)

- Implemented attention from scratch for German to English, generating BLEU-4 of 13.5 with 10k samples.
- Improved decoding with a lexical module, achieving a 2.6-point BLEU gain over the greedy decoding.

### 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.