

# Shobhit Maheshwari

96 Holyrood Rd, Edinburgh EH88FH | +44-7826679948 | [shobhit.workds@gmail.com](mailto:shobhit.workds@gmail.com) | [Linkedin](#) | [Github](#)

## SUMMARY

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.

## EDUCATION

### University of Edinburgh

MSc in Data Science (*Distinction in 90/120 credits*)

Edinburgh, UK

Sept. 2023 – Present

### University of Delhi

B. Tech. in Information Technology and Mathematics (*89.5/100*)

Delhi, India

Aug. 2015 – June 2019

## EXPERIENCE

### Senior Data Scientist

Jan. 2022 – Aug. 2023

Roadzen

New Delhi, India

- 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

June 2020 – Dec. 2021

Roadzen

New Delhi, India

- 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.

### Data Scientist

July 2019 – June 2020

Spoonshot

Bangalore, India

- Developed ingredient embeddings using weighted DeepWalk; Used nodes as ingredients and the edge traversal probability with a heuristic of flavour pairing theory and co-occurrence of ingredients in a product.
- Handled 100 million ingredient combinations using a PySpark pipeline for edge weights generation.
- Extracted and standardised the nutrition panel information from products using Fast-RCNN and Azure OCR.

## PROJECTS

### QA framework for long financial PDFs | *Python, PyTorch, LLM, Vector DB, RAG, Langchain, MapReduce*

- Historical annual reports scraped for FTSE all-share accounting for 70GB data.
- Implemented ranked MapReduce to achieve 16% improvement on QA with equivalent processing time as RAG.

### Neural Machine Translation | *Python, PyTorch, LSTM, Attention, Transformers*

- Multi-headed attention implemented 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.

### Image Captioning | *Python, Tensorflow, Keras, CNN, LSTM, Transfer Learning*

- 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.

## TECHNICAL SKILLS

**Languages:** Python, SQL, HTML, CSS, C/C++

**ML, NLP, Computer Vision:** Regression, Dimensionality reduction, Random Forest, XGBoost, Adaboost, Clustering, MLP, RNN, LSTM, Attention, Transformers, LLM, BERT, RAG, Vector DB, CNN, GAN

**Libraries:** NLTK, Spacy, Pandas, Matplotlib, Numpy, Scikit-Learn, Keras, TensorFlow, PyTorch, PySpark, LangChain

**Frameworks:** Selenium, Scrapy, Flask, FastAPI

**Developer Tools:** Git, Docker, Kubernetes, Nginx, AWS, Airflow, Jenkins, VS Code