Shobhit Maheshwari
Data Scientist, Machine Learning Engineer

87 Saughton Mains St, Edinburgh EH11 3HB | +44-7826679948 | shobhit.workds@gmail.com | Portfolio | Linkedin | Github

Data Scientist with a strong foundation in Mathematics and 4+ years of work experience in Computer Vision and Natural Language Processing (NLP). Experienced in Python, SQL with proficiency in Machine Learning libraries like PyTorch and TensorFlow. Skilled in deploying AI solutions, optimizing ML workflows, and collaborating with cross-functional teams.

EDUCATION

- **University of Edinburgh**   Edinburgh, UK
  M. Sc. In Data Science (Distinction grades in 90/120 credits)   *Sep 2023 – Present*

- **University of Delhi**   Delhi, India
  B. Tech. in Information Technology and Mathematics (89.5)   *Aug 2015 – Jun 2019*

EXPERIENCE

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

  - Streamlined ML workflow from data ingestion to model serving using FastAPI, Airflow, and Jenkins, resulting in continuous training and efficient deployment of the inhouse 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.
  - Enhanced the 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.
  - 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 weights generation.
  - Implemented Fast-RCNN to extract nutrition panel with a mAP of 85 from product images.
  - Used Azure OCR to extract nutrition component and its corresponding value with an accuracy of 93%.

PROJECTS

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

  - Contrasted GPT with 4-bit quantised Llama and Mistral on QA tasks; GPT outperformed both by ~12%.
  - Achieved a 16% improvement over RAG using a ranked MapReduce pipeline without additional processing time.

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

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

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

TECHNICAL SKILLS

Python, SQL, Pandas, Matplotlib, Numpy, OpenCV, Scikit-Learn, Keras, TensorFlow, PyTorch, PySpark, LangChain, NLTK, Spacy, Selenium, Scrapy, Flask, FastAPI, Git, Docker, Kubernetes, Nginx, AWS, Airflow, Jenkins