# Neetu Kushwaha

+44-798-440-0316 | neetumits@gmail.com | https://neetu-kushwaha.github.io/

https://www.linkedin.com/in/neetu-kushwaha-71460851/ | https://github.com/Neetu-Kushwaha |

Oxford, United Kingdom

## Research Interests

Natural Language Processing (NLP) with Transformer and BERT models, Deep Learning, Neural Network Debugging, Object Detection in Satellite Imagery, Geospatial Data Analysis, Machine Learning, Metaheuristic Algorithms

## Skills

- **Data Science & Machine Learning:** PyTorch, TensorFlow, NumPy, Scikit-learn, Spacy, NLTK, LlamaIndex, LangChain, LangGraph, Agentic AI, Transformers (BERT, GPT, LLaMA), Knowledge Graph
- **Programming Languages:** Python, MATLAB, JavaScript, C++
- **Web Technologies:** Flask, Selenium, BeautifulSoup
- **Database Systems:** MySQL, Neo4j Graph Database
- **Geospatial Analysis:** GeoPandas, Shapely, QGIS
- **Cloud Technologies:** Google Cloud Platform (GCP), AWS
- **DevOps & Version Control:** Git, Docker

## Experience

- **Oxford Sustainable Finance Group, SSEE, University of Oxford [🌐]** — *Oct 2024 - Present*
  *Postdoctoral Research Associate in Machine Learning* — Oxford, United Kingdom
  - **Mapping Asset Ownership with NLP and Computer Vision**
- **The Alan Turing Institute[🌐]** — *July 2022 - May 2024*
  *Postdoctoral Research Associate in Machine Learning* — London, United Kingdom
  - **Machine learning in sustainable finance**
- **INRIA [🌐]** — *March 2020 - Feb 2021*
  *Postdoctoral Researcher* — Rennes,France
  - **Pattern Mining for Neural Networks Debugging**

## Education

- **Indian Institute of Technology Roorkee** — *December 2014 - October 2019*
  *Ph.D. in Machine Learning* — Roorkee, India
  - Thesis: Enhanced Metaheuristics Based Clustering Algorithms and Their Applications.
- **Dr B R Ambedkar National Institute of Technology, Jalandhar** — *August 2010 - July 2012*
  *M.Tech in Computer Science & Engineering* — Jalandhar, India
  - Thesis: Software Cost Estimation using Soft Computing Method
- **Madhav Institute of Technology & Science** — *August 2005 - July 2009*
  *B.Tech. in Computer Science & Engineering* — Gwalior, India

## Projects

- **Knowledge RAG System (In Development)**
  *Tools: Python, Pandas, LangChain, Neo4j, open source LLM-LLama3.2*
  - Developing a knowledge-driven RAG system using LangChain and Neo4j for efficient retrieval and organization of nature finance-specific information.
- **RAG for Data Center Information Extraction**
  *Tools: Python, Pandas, LangChain, open source LLMs, pdfplumber-llm*
  - Developed a RAG system using LangChain to extract and structure data center information from open-source PDFs.
  - Utilized pdfplumber-llm for text extraction, implemented chunking strategies, and integrated LLaMA 3.3 for accurate question answering.
- **Brick Kiln Detection Using Deep Learning**
  *Tools: Python, YOLOv8, PyTorch, Sentinel-2, Google Maps API*
  - **External Contributor, APAD - Air Pollution Asset-Level Detection:** Contributed to large-scale brick kiln detection using deep learning and satellite imagery.
    * Developed a two-phase pipeline integrating a Random Forest classifier on Sentinel-2 imagery for initial detection and YOLOv8 for high-resolution object detection and classification.

    - Enhanced accuracy and scalability while reducing reliance on high-resolution imagery, enabling improved air pollution monitoring.

- **GeoAsset Project**
  *Tools: Python, Selenium, Spacy, NLTK, s-BERT, PyTorch*
  - **Asset Ownership:** Mapped asset-level data to companies using NLP and computer vision, extracting information from various sources and satellite images.
    - Applied NLP techniques, including Named Entity Recognition (NER), for data extraction, analysis, preprocessing, and duplicate removal, and developed a ranking system to rank relevant information from raw text.
    - Developed a RAG system for question answering, leveraging NLP models for efficient information retrieval and generation of answers.
    - Leveraged satellite imagery for precise asset identification through image classifi- cation, followed by image segmentation to accurately separate production sites.

- **Decarbonizing Agriculture - Barclay Project**
  *Tools: Python, Pandas, GeoPandas, Selenium, QGIS*
  - **Decarbonising Agriculture:** This project aims to reduce carbon emissions in UK agriculture by measuring farm-level greenhouse gases and developing decarbonization plans in collaboration with B4ICA.
    - Conducted end-to-end data wrangling, including scraping, matching, preprocessing, and duplicate removal of diverse sources.
    - Leveraged advanced NLP methods, particularly s-BERT, to enhance data quality, with a focus on accurate entity resolution for name matching.

- **Precision Refinement in CNNs: Faulty Neuron Identification**
  *Tools: PyTorch, Numpy, Scikit-learn, Keras*
  - Developed targeted debugging strategies for isolating and addressing faulty neurons in CNNs.

## PUBLICATIONS

C=CONFERENCE, J=JOURNAL, S=IN SUBMISSION

**[S.1]** Hamdani, Muhammad Suleman Ali, Khizer Zakir, Neetu Kushwaha, Syeda Eman Fatima, and Hassan Aftab Sheikh. "Brick Kiln Dataset for Pakistan's IGP Region Using AI.", Dataset available at: Zenodo Record 14038648.

**[J.1]** Kushwaha, Neetu, and Millie Pant."Fuzzy electromagnetic optimisation clustering algorithm for collaborative filtering." Journal of Experimental & Theoretical Artificial Intelligence 33.4 (2021): 601-616.

**[J.1]** Kushwaha, Neetu, and Millie Pant."Textual data dimensionality reduction-a deep learning approach." Multimedia Tools and Applications 79.15 (2020): 11039-11050.

**[J.1]** Kushwaha, Neetu, and Millie Pant."Link based BPSO for feature selection in big data text clustering.", Future generation computer systems 82 (2018): 190-199.

Note: For complete list, please see here.

## WORKSHOPS

- **Organized member of first workshop on NLP for Climate : ClimateNLP 2024 (ACL 2024)** *[🌐]*