# Ankon Bhowmick

ankonbh@gmail.com, +919667055306, https://www.linkedin.com/in/ankon-bhowmick-2a8805193/

## Personal Summary

Aspiring Data Scientist with a strong foundation in statistics, machine learning, and data analysis. As a recent graduate with an MSc in Data Science and Analytics at the University of Leeds and a BTech in Mathematics and Computing from Delhi Technological University, I bring expertise in exploratory data analysis and predictive modelling, alongside being skilled in Python, R, and SQL. My analytical rigor and commitment to continuous learning equip me to deliver impactful insights and innovative solutions.

## Education

1. **MSc Data Science and Analytics**, University of Leeds, (September 2023 - December 2024), 2:1 (UK)
2. **BTech Mathematics and Computing**, Delhi Technological University, (August 2019 – June 2023), CGPA – 8.33/10

## Internships

1. **PwC**, Gurgaon, India, (June 2022 - August 2022)
   Intern/Trainee, Advisory- Technology Consulting
   - Completed company training programs to gain a basic understanding in SAP ERP technology.
   - Collaborated with manager to perform an in-depth analysis of a substantial data dump containing over 50,000 SAP incident reports from a client company.
   - Implemented advanced data analysis techniques and Python programming to explore the data which contained 30 features, identifying both positive trends and potential areas of concern.
   - Deployed a decision tree model to predict the possibility of successful resolution of incidents with an accuracy of 97.7%. Modularised model code for ease of use by other team members.

## Certifications

1. **Data Science Project-Based Learning and Internship Program Using R, Python Programming, And Tableau**, StepUp Analytics, Online, (December 2020 - February 2021)
2. **Microsoft Excel (Office 2019)**: Exam Reference #45320081

## Skills

1. **Programming Languages**: Python, R, SQL
2. **Python Libraries**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Transformers, OpenCV, PyTorch, TensorFlow
3. **R Libraries**: dplyr, ggplot2, caret, glm, rpart, timeSeries, rstanarm
4. **ML Algorithms**: Linear Regression, Logistic Regression, KNN, Decision Trees, Naïve Bayes, K-Means, DBSCAN, Random Forest, PCA, Multi-level Regression
5. **Deep Learning**: CNN, RNN, LSTM
6. **NLP**: Text Preprocessing, Embeddings, Vector Database (FAISS), BERT (HuggingFace), GenAI (LangChain)
7. **Other Relevant Coursework**: Time Series, Data Structures & Algorithms, Digital Image Processing

## Academic Projects

1. **Predicting the Course of the Russo-Ukrainian War**
   - Built a BERT-based NLP model to classify war-intensity from 32,000+ news articles.
   - Keywords: Python, Web Scraping, Selenium, Beautiful Soup, NLP, BERT, Hugging Face Transformers, PyTorch, Tensor Datasets, Text Classification, Data Preprocessing
2. **Image Classification and Captioning with CNNs & RNNs**
   - Developed CNN and RNN models for image classification and captioning, tackling overfitting with data augmentation and transfer learning to improve performance.
   - Keywords: Python, TensorFlow, PyTorch, CNN, RNN, Neural Networks, Image Classification, Transfer Learning, Data Augmentation, Hyperparameter Tuning, Model Optimisation
3. **Portfolio Query Assistant using RAG**
   - Built a retrieval-augmented generation (RAG) system using LangChain to process and query Word and PDF documents with semantic search powered by FAISS and Google GenAI.
   - LangChain, RAG, Document Loaders, Google GenAI, RecursiveCharacterTextSplitter, FAISS, Semantic Search, Vector Database, Python
4. **Estimating Public Opinion on Gun Legislation using MRP:**
   - Trained hierarchical models on over 60,000 samples, incorporating individual- and group-level predictors, and validated performance using multiple statistical metrics.
   - Keywords: MRP, rstanarm, R, Python, Pandas, NumPy, ggplot2, matplotlib, MAE, WAIC, hierarchical modelling, poststratification, public opinion estimation.