## Vandana Sharma

Data scientist & AI Engg. | SDE | Project Management

Aspiring Data Scientist with hands-on experience in AI/ML, specializing in complex models like RNNs, LSTMs, Transformers, Generative AI, and LLMs. Drives business impact through sophisticated ML pipelines.

## Portfolio | GitHub | LinkedIn

+91 7348482348

vandana.sharma727272@gmail.com

## **EXPERIENCE**

**Data Science Intern (Virtual Program)** 

Boston Consulting Group (BCG) - Forage | Feb 2025 - May 2025

Completed real-world business case modules simulating data science tasks at BCG, including data exploration, data visualization using Python and Excel, and delivering actionable client insights based on statistical analysis.

**Machine Learning Intern** 

Ekak Innovation Pvt. Ltd., Kanpur | Dec 2024 – Jan 2025

Optimized machine learning models for production-level projects, resulting in a **15% improvement** in prediction accuracy using Scikit-learn. Automated end-to-end data preprocessing pipelines for large-scale datasets using Pandas and Scikit-learn, enhancing workflow efficiency and model readiness.

**Data Science Intern** 

Earth5R (Remote) | Jun 2024 - Sep 2024

Built HR analytics dashboard in Tableau, reducing manual reporting time by 40% and enabling leadership to make data-driven decisions.

Developed a Python-based CLI application to automate reporting for non-technical stakeholders.

**Data Science Intern** 

YBI Foundation (Remote) | Jan 2024 - Feb 2024

Developed a spam email classification model using Scikit-learn, achieving 98% accuracy.

## **EDUCATION**

B.Tech - Computer Science & Engineering

Chhatrapati Shahu Ji Maharaj University, Kanpur (2021 – 2025)

**PROIECTS** 

Vehicle Insurance (MLOps)

(Ongoing) **GitHub** 

Twitter Sentiment Analysis with Deep Learning (LSTM)

Developed a Bidirectional LSTM for Twitter sentiment analysis, achieving 99% test accuracy and near-perfect precision/recall, applying robust NLP preprocessing and early stopping in TensorFlow/Keras.

House Price Prediction System (MLOps)

Built an industrial end-to-end ML pipeline (ZenML, DVC, MLflow) with CI/CD automation.

Olympics Performance Analysis Web App Live Web App | GitHub

Interactive Streamlit dashboard with medal, country, and athlete insights (1896-2016).

HR Analytics Dashboard using Tableau

Visualized attrition and salary KPIs; enabled leadership decisions via interactive dashboard.

**Brain Tumor Classification with CNN** 

GitHub

Achieved 82% accuracy using VGG19 and data augmentation techniques.

AI Gym Tracker & Pose Detector

Built real-time pose detection app with 85% rep-count accuracy using MediaPipe.

**SKILLS** 

Programming: Python, SQL, Bash,

Linux Shell

Machine Learning & Data Science: EDA, Classification, Regression, Clustering, PCA, Supervised & Unsupervised Learning

Deep Learning: ANN, CNN, RNN, LSTM, GRU, Transformers, Generative AI, LLMs

NLP: Text Preprocessing, Lemmatization, Tokenization, Sentiment Analysis, Text Classification

**Tools** & Frameworks: Scikit-learn, TensorFlow, Keras, NumPy, Pandas, NLTK, OpenCV

MLOps: MLflow, DVC, TFX, Git, Jenkins, CI/CD Pipelines

Visualization: Tableau, Power BI, Streamlit, Plotly, Matplotlib, Seaborn

Cloud & Deployment: AWS (EC2 & S3), GCP,

Docker, Flask, Netlify, Render, Azure Databases: MySQL, SQLite, MongoDB

**CERTIFICATIONS** 

LLMs (Large Language Models) -Google Cloud (Ongoing)

Data Science Cisco Networking Academy (2024)

Machine Learning - Simplilearn (2024) Python Programming – Great Learning (2023)

**LANGUAGES** 

English, Hindi