Harsh Dubey

MCA (AI/ML) Student — Aspiring Machine Learning Engineer
Lucknow, Uttar Pradesh, India
harsh44dubey@gmail.com | linkedin.com/in/iamharshdubey | github.com/iamharshdubey

Professional Summary

Results-driven MCA student specializing in Artificial Intelligence and Machine Learning with hands-on experience building production-ready ML models. Proficient in Python, scikit-learn, NLP, and end-to-end ML pipeline development. Proven ability to solve complex problems through data-driven solutions and deploy scalable ML applications. Seeking ML Engineering or Data Science opportunities to contribute to cutting-edge AI initiatives.

Education

Chandigarh University — Uttar Pradesh, India
Master of Computer Applications (MCA) – Artificial Intelligence & Machine Learning — *Aug 2025 – Aug 2027*

- Specialization: AI/ML with focus on Deep Learning, Neural Networks, and Advanced Data Analytics
- Relevant Coursework: Machine Learning, Deep Learning, NLP, Computer Vision, Statistical Analysis

University of Lucknow — Lucknow, India
Bachelor of Computer Applications (BCA) – Computer Science — *Graduated 2022*

- Foundation in Data Structures, Algorithms, Database Management, and Software Engineering

Technical Skills

Programming Languages: Python, SQL, JavaScript, HTML/CSS
Machine Learning: Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, Random Forest, SVM, Neural Networks
Data Science & Analytics: Pandas, NumPy, Matplotlib, Seaborn, Data Preprocessing, Feature Engineering, EDA
Natural Language Processing: NLTK, spaCy, Text Classification, Sentiment Analysis, TF-IDF, Word2Vec
Tools & Technologies: Git/GitHub, Jupyter Notebook, VS Code, Google Colab, Docker, REST APIs
Databases: MySQL, PostgreSQL, MongoDB
Cloud Platforms: AWS, Google Cloud Platform (GCP), Azure (Basic)
Soft Skills: Problem-Solving, Team Collaboration, Agile Methodologies, Technical Documentation

Projects

Fake Job Detector | *Python, NLP, Scikit-learn, Machine Learning* — GitHub

- Developed an end-to-end ML classification system to detect fraudulent job postings with 94% accuracy using NLP techniques and ensemble learning methods
- Implemented comprehensive text preprocessing pipeline including tokenization, lemmatization, and TF-IDF vectorization on dataset of 18,000+ job listings
- Engineered features from job descriptions, requirements, and company profiles to improve model performance
- Deployed model with REST API for real-time predictions and created interactive dashboard for results visualization
- Technologies: Python, NLTK, Scikit-learn, Pandas, Random Forest, Logistic Regression, Flask

Customer Churn Prediction System | *Python, Machine Learning, Data Analytics* — GitHub

- Built predictive model to identify at-risk customers with 89% precision using ensemble methods and hyperparameter tuning
- Performed extensive EDA and feature engineering on 10,000+ customer records to extract meaningful patterns
- Implemented cross-validation and A/B testing framework to validate model performance across different segments
- Technologies: Python, Pandas, XGBoost, Matplotlib, Seaborn, Jupyter Notebook

Sentiment Analysis Engine | *NLP, Deep Learning, Python* — GitHub

- Created sentiment classification model using LSTM networks achieving 91% accuracy on product reviews dataset
- Preprocessed and cleaned 50,000+ text samples using advanced NLP techniques and word embeddings
- Implemented model comparison framework testing Naive Bayes, SVM, and Deep Learning approaches
- Technologies: Python, TensorFlow, Keras, NLTK, Word2Vec, spaCy

Achievements & Certifications

Machine Learning Specialization Coursera / Stanford University

- Completed comprehensive ML course covering supervised/unsupervised learning and neural networks

Open Source Contributions GitHub

- Active contributor to ML and data science open-source projects with consistent GitHub activity
- Maintained portfolio of 10+ end-to-end machine learning projects demonstrating practical implementation skills

Hackathon Participation Multiple Events

- Participated in ML/AI hackathons solving real-world problems using data-driven approaches

Additional Information

Research Interests: Large Language Models (LLMs), Computer Vision, Reinforcement Learning, MLOps
Languages: English (Fluent), Hindi (Native)
Professional Development: Continuously learning through online courses, research papers, and hands-on projects