VARNIT KUMAR

kumar.varnit.16@gmail.com | +91-8287848299 | GitHub | LinkedIn | Delhi, India

PROFESSIONAL SUMMARY

Aspiring ML Engineer with 6 months of hands-on industry experience at Amazon, specializing in data analysis, machine learning model development, and end-to-end ML pipelines. Proven track record of building production-ready ML solutions achieving 98% accuracy. Proficient in Python, TensorFlow, scikit-learn, and cloud deployment (AWS, GCP). Passionate about solving real-world AI problems and eager to contribute to practical projects while gaining mentorship in a growth-oriented environment.

EDUCATION

Master of Computer Applications (MCA) *September 2024 – September 2026*

BCIIT & Innovation, Guru Gobind Singh Indraprastha University, Delhi **CGPA: 9.2/10**

Relevant Coursework: Machine Learning, Deep Learning, Data Structures & Algorithms, Statistical Analysis, Database Management

Bachelor of Computer Applications (BCA) *July 2020 – July 2023*

Institute of Technology & Management, Hemvati Nandan Bahuguna Garhwal University **CGPA: 7.2/10**

TECHNICAL SKILLS

Programming Languages:	Python (pandas, NumPy, scikit-learn, TensorFlow, Keras, NLTK), SQL, R
ML/AI Frameworks:	TensorFlow, Keras, Scikit-learn, XGBoost, Random Forest, SVM, CNN, NLP, Logistic Regression
Tools & Platforms:	AWS (S3, Lambda, SageMaker), GCP, MLflow, Power BI, Tableau, Git, Flask, ETL Tools
Databases:	MySQL, MongoDB, PostgreSQL
Core Competencies:	Data Analysis, Feature Engineering, Model Deployment, EDA, Statistical Analysis, Data Visualization

WORK EXPERIENCE

ML Data Associate *June 2024 – December 2024*

Amazon, Gurugram, India

- Analyzed large-scale datasets (500K+ records) to identify data quality trends and improve internal labeling accuracy by **25%**, directly enhancing machine learning model training quality and performance
- Collaborated with data scientists on 3 classification ML models; contributed to data preprocessing, feature engineering, and pipeline optimization that reduced manual annotation errors by **30%**
- Designed interactive dashboards using **Power BI** and **Matplotlib** for cross-functional stakeholders, enabling faster data-driven decision-making; optimized AWS ETL pipelines improving processing speed by 15%
- Participated in Agile sprints with **95% on-time delivery rate**; recognized by team leads for analytical contributions and data visualization efficiency

KEY PROJECTS

Android Malware Detection System *Python, Scikit-learn, XGBoost, SHAP*

- Built a machine learning classification model to detect malicious Android apps using static code analysis on **10,000+ APK samples**; achieved **98% accuracy** through hyperparameter tuning with Random Forest and XGBoost
- Performed comprehensive feature extraction from app permissions, API calls, and code signatures; optimized model performance using grid search and cross-validation techniques
- Implemented SHAP (SHapley Additive exPlanations) for model interpretability, visualizing feature importance to provide transparency and actionable insights for security teams

Farm-IQ: AI-Powered Smart Farming Assistant *Python, Flask, TensorFlow, ML*

- Developed full-stack web application providing crop recommendations, fertilizer guidance, and plant disease detection using Python, Flask, and machine learning models
- Integrated real-time weather API data with Random Forest classifier achieving **92% accuracy** for crop recommendations based on soil parameters and climate conditions
- Implemented CNN-based image classification model using TensorFlow/Keras for plant disease detection, trained on **10,000+ labeled crop images**; deployed Flask REST API on cloud platform for production use

E-Commerce Sentiment Analysis *Python, NLP, Scikit-learn*

- Enhanced product review sentiment classification achieving **85-90% accuracy** using natural language processing and machine learning techniques on **50,000+ customer reviews**
- Implemented comprehensive text preprocessing (cleaning, tokenization, stopword removal, lemmatization) and TF-IDF vectorization with multiple ML models: Logistic Regression, SVM, and Random Forest

- Performed comparative model analysis to identify optimal algorithm; generated insights on customer satisfaction patterns for e-commerce platforms

Traffic Sign Detection System *TensorFlow, OpenCV, CNN*

- Developed CNN-based traffic sign classification model using TensorFlow and OpenCV on German Traffic Sign Recognition Benchmark dataset with **40+ sign classes**; achieved **94% accuracy**
- Applied data augmentation techniques (rotation, zoom, brightness adjustment) to improve model robustness and generalization capabilities
- Implemented real-time video stream processing for autonomous driving applications using OpenCV for image preprocessing and model inference

CERTIFICATIONS & TRAINING

Professional Certifications: Data Scientist Professional (DataCamp) • Data Analyst Associate (DataCamp) • Python Professional Certificate (OpenEDG Python Institute)
Specialized Training: Complete SQL Bootcamp (Udemy) • Machine Learning & AI Fundamentals (AWS) • Essentials of Prompt Engineering (AWS) • Google Analytics Certification • SQL Certification (HackerRank)
Industry Simulations: Deloitte Australia, Accenture North America, Tata, and Quantium Data Analytics Job Simulations (Forage, 2024-25)

LEADERSHIP & ACHIEVEMENTS

Cultural & Engagement Ambassador, Amazon (June 2024 – December 2024)
Led cultural engagement initiatives, organized virtual and in-person team events, and acted as communication bridge between leadership and peers; recognized by team leads for boosting team morale and improving cross-team collaboration

Cricket Team Captain, BCIIT (2024 – Present)
Leading university cricket team in inter-college tournaments with responsibility for team selection, strategy development, and performance management