HARSH VERMA

harshkrish2323@gmail.com | +91 8928806258 | LinkedIn | GitHub | LeetCode
Mumbai, India | **Available January – June 2027 for a 6-month Data Science / ML internship** (remote or on-site)

Education

Thakur College of Engineering and Technology (TCET), University of Mumbai **July 2023 – 2027 (Expected)**
Bachelor of Engineering in Information Technology, ***CGPA: 9.25/10*** *Mumbai, Maharashtra*

- **Coursework:** Probability & Statistics, Linear Algebra, Calculus, Machine Learning, Deep Learning, Data Mining & Warehousing, Big Data Analytics, DBMS, DSA

Projects

Customer Churn Prediction (BCG X Job Simulation) | *Python, Pandas, Scikit-learn* | Certificate **May – Sep 2026**

- Explored 14,606 SME customers and 193,002 monthly price records; engineered 20 features (tenure, contract timing, Dec–Jan price changes, peak/off-peak spreads); found price sensitivity a weak churn driver.
- Class-weighted (1:6) a 500-tree Random Forest: on a stratified 3,652-customer test set, churn recall rose from 8.2% to 43.7% at 22% precision, a 2.3× lift over the 9.7% base rate (ROC-AUC 0.698).
- Executive summary: price sensitivity was a weak driver; recommended retention targeting on consumption and margin signals.

Loan Approval Prediction | *Python, Scikit-learn, XGBoost* | GitHub **Aug 2026**

- Built banking features (loan-to-income, EMI, income per dependent) with log transforms on 614 applicants and compared 8 classifiers: best hold-out accuracy 90.3%, F1 0.933 (stratified, n = 62).
- Showed 7 of 8 models gave identical predictions driven by credit history and differed only in ranking (ROC-AUC 0.85–0.91), so chose interpretable logistic regression.

Dialogue Text Summarizer | *PyTorch, Hugging Face Transformers, T5* | GitHub **Oct 2026**

- Fine-tuned T5-small for abstractive summarization of multi-speaker SAMSum chats (4,000-dialogue subset, 512/150-token inputs/targets, 6 epochs on a Colab T4 GPU); validation loss fell from 0.380 to 0.350.
- Built regex cleaning and tokenization, 4-beam search decoding, and a small FastAPI endpoint serving the saved model.

ML Algorithms from Scratch | *NumPy* | GitHub **2026**

- Implemented linear regression (batch gradient descent and the normal equation), logistic regression (sigmoid with log-loss gradients) and KNN classification/regression without ML libraries.

Experience

TechFlex Solutions — Network Engineer Intern (on-site) **Dec 2025 – Jan 2026**
Deployed at client Finacus (fintech): routing, virtualization, log/traffic monitoring for anomalies *Mumbai*

Beta Microsoft Learn Student Ambassador (MLSA) **July 2024 – June 2025**
Student Ambassador, Microsoft Learn Student Ambassadors program

- Facilitated community activities for 100+ participants, building technical literacy in cloud, AI and programming foundations.
- Developed 3 hands-on sessions: Azure AI services, deploying a resume website with GitHub, and a no-code Azure AI chatbot.

Technical Skills

Machine Learning: Regression, classification, tree ensembles, class imbalance, model evaluation, transformer fine-tuning
Math & Statistics: Probability, statistical inference, hypothesis testing, linear algebra, calculus, optimization
Languages & Libraries: Python, SQL, C++ (DSA); NumPy, Pandas, Scikit-learn, XGBoost, Matplotlib, PyTorch, Hugging Face
Tools & Platforms: Jupyter, Google Colab, Git/GitHub, Microsoft Azure (Azure ML), MySQL, MongoDB, Linux (Ubuntu)

Certifications

Microsoft Certified: Machine Learning Operations Engineer Associate | AI-300 Exam **October 2026**
Microsoft Certified: Azure AI Fundamentals | AI-900 Exam **November 2025**

Achievements

- **GATE 2026 (CS):** Score **419**, AIR **15,677** (top 7.5% of ∼2.1 lakh), qualified in 3rd year; preparing for GATE DA 2027.
- **Naukri Campus Young Turks 2026:** 93.35 percentile in Round 1 of the national skill contest.
- **LeetCode:** 200+ problems solved (73 medium, 6 hard); 140-day streak in 2025.