Katta Manoj Kumar

+91-6281215816 | — kattamanojkumar2005@gmail.com | — LinkedIn | — GitHub

EDUCATION

- **Vellore Institute of Technology, Andhra Pradesh** *2022-26*
  B.Tech - Computer Science and Engineering CGPA: 9.06
- **Narayana Junior College** *2020-22*
  M. P. C, Board of Intermediate Education, A. P. CGPA: 9.61
- **Bhashyam High School** *2020*
  State Board of Secondary Education Percentage: 100%

EXPERIENCE

- **Research Internship -IIT Hyderabad** (May 2025-July 2025)
  - Developed a custom static analysis pipeline by extending Java bytecode frameworks (ASM, WALA) to extract control flow graphs , call graphs, and other program structures from large-scale datasets like Java250 (75k+ files) and BigCloneBench.
  - Automated the end-to-end preprocessing workflow using optimized python scripts, significantly reducing data extraction time and enabling scalable downstream ML tasks.
  - Contributed to building a reusable research-grade infrastructure for code representation learning, enhancing dataset generation efficiency and adaptability for various program analysis objectives.

- **Amazon ML Summer School 2024** (June 2025-Aug 2025)
  - Among the top 3,000 selected from over 200,000 applicants for Amazon ML Summer School 2024, a highly competitive program focused on advanced Machine Learning concepts.
  - Covered Supervised Learning, Deep Neural Networks, Dimensionality Reduction, Unsupervised Learning, Reinforcement Learning, Generative AI(LLMs), and Causal Inference.
  - Engaged in interactive sessions with Amazon Scientists, learning about industry applications of ML.

PROJECTS

- **Byte2Vec – Custom Embeddings for Bytecode**
  - Engineered a custom bytecode embedding pipeline by disassembling Python files into low-level instruction streams and assigning consistent numeric indices across tokens, inspired by VexIR2Vec.
  - Generated structured triplet data and trained 200-dimensional opcode embeddings using OpenKE's translational models, capturing semantic relationships among operations.
  - Achieved 95%+ Accuracy in opcode analogy prediction and clustered functionally similar opcodes using KMeans, validating the quality and generalizability of the learned vector space.
- **Code Clone Detection using Siamese Neural Networks**
  - Leveraged custom embeddings on BigCloneBench to train a Siamese Neural Network for identifying semantic and near-miss code clones across large Java Source code pairs.
  - Designed a robust twin-network architecture to compare function-level vector representations, achieving 96% classification accuracy on clone detection tasks.
  - Optimized model performance with effective negative sampling, batch tuning, and inference pipeline enhancements for scalable static code analysis.

TECHNICAL SKILLS AND INTERESTS

Languages: Java, Python, R, C.
Libraries and Frameworks : NumPy, Pandas, Matplotlib, Spring, SpringBoot, PyTorch, Wala, ASM .
Databases: MySQL, PostgreSQL, Oracle.
Software development tools: Git, GitHub, GitLab.
Relevant Coursework: Data Structures & Algorithms, Operating Systems, OOPS, Database Management System, Cloud Computing, Natural Language Processing, Deep Learning.
Soft Skills: Leadership, Communication, Client relations, Learning agility.

ACHIEVEMENTS

- **Solved over 600 problems on Leetcode and max contest rating 1630.**
- **Amazon Machine Learning Summer School Program** *July 2024*
- **3rd/380 Teams, University Engineering Clinics Competition,** VIT-AP. *August 2023*
- Achieved a top 1% ranking in AP EAPCET (Rank 2000) and TS EAMCET (Rank 1500) among 200000 candidates each, demonstrating competitive proficiency in state level entrance exams.

EXTRACURRICULAR ACTIVITIES

- Actively engaged in outdoor games consistently demonstrating teamwork, strategy, and physical endurance
- Successfully organized a large-scale cultural festival at VIT, overseeing 20+ events and engaging over 500 participants. Managed planning logistics and budgeting while leading a cross-functional team to deliver a seamless experience, showcasing leadership and strategic coordination.