Harsh Mishra

mishraharsh169@gmail.com | linkedin.com/in/harsh-mishra-515624144/ | harshm16.github.io/

Summary

Machine Learning Engineer with extensive industry and academic research experience in applied machine learning. Skilled in integrating advanced AI technologies, automating infrastructure and data engineering workflows.

Technical Skills

Proficient: Python, R, C, SQL, Bash scripting
Machine Learning Libraries/Framework: Pytorch, TensorFlow, Scikit-learn, Hugging Face, Langchain
Tools/Software: Tableau, Docker, AWS (S3, Lambda, EMR, EBS, SageMaker, Kinesis), Jira, Latex, Postman, Git, Kubernetes, CromaDB

Professional Experience

AI & Machine Learning Engineer, Aston Villa - Birmingham, UK **Dec. 2024 – Present**
Technologies Used: Python, LLMs, Transformers, REST APIs

- **Developing ML models:** Developing end to end ML models using event and tracking data to generate insights for men's and women's teams, supporting player recruitment, scouting and tactical decision-making.

Machine Learning Engineer, KnetMiner - Harpenden, UK **Feb. 2024 – Dec. 2024**
Technologies Used: Python, Neo4j, Langchain, OpenAI, Hugging Face Transformers, REST APIs

- **Developed RAG solutions:** Led the integration of advanced AI techniques into existing knowledge graphs by building Retrieval-Augmented Generation (RAG) solutions on top of them, improving data retrieval and knowledge discovery processes.
- **Developed a Chatbot interface:** Created an in-house AI-integrated chatbot to serve as a user-friendly interface for querying and interacting with knowledge graphs, streamlining user access to information.

Software Engineer, Hewlett Packard Enterprise (HPE) - Bangalore, India **Jan. 2019 – July 2021**
Technologies Used: Python, Bash, REST API, Ansible, Docker, Jenkins, Neo4j, Cypher, SQL, Kubernetes

- **Automated MLOps Deployment**: Developed scripts to automate the deployment of MLOps as a Service, utilizing REST APIs to replicate deprecated PowerShell scripts in Python and leveraging Ansible for automated deployment, enhancing operational efficiency.
- **Simulated Cyber Threat Patterns:** Simulated cyber threat patterns using graph databases in Neo4j and wrote complex Cypher and SQL queries to detect such patterns, serving as a benchmark to evaluate and test server solutions offered for security purposes.
- **Built and Deployed DataOps Pipeline:** Developed and deployed a robust DataOps pipeline using open-source applications and authored bash scripts to automate the CI/CD pipeline, significantly reducing deployment times and increasing reliability in data processing workflows.

Academic Research Experience

Researcher, Computer Science Department, University of Illinois at Chicago **Sep 2021 – Dec. 2023**
Technologies Used: Python, Pytorch, TensorFlow, Docker, Linux, Latex

- **Trained Score-Based Generative Models:** Conducted research on training score-based generative models using non-Gaussian noise. The paper was selected for a poster presentation at the MMLS 2023 conference. Preprint available on arxiv.org/pdf/2302.02336.pdf. My Masters's Thesis on this topic can be found on Indigo-UIC: https://doi.org/10.25417/uic.23661801.v1.
- **Algorithm Development for Kernel Methods:** Developed an algorithm to convert categorical labels/features to continuous labels, enabling the use of kernel methods for node classification and other Graph Neural Network (GNN) tasks. The algorithm and experiments on Event Stream and Pose Estimation data are available on github.com/harshm16/GNN.
- **Optimization-Based Training in Distributed ML Systems**: Co-authored a paper on using optimization-based training methods designed to enhance distributed Machine Learning systems, specifically addressing Byzantine worker issues. This work was selected for presentation at ICLR 2024. Print available on openreview.net/forum?id=7avlrpzWqo.

Education

University of Illinois at Chicago (UIC), USA **Aug 2021-May 2023**
Master of Science, Computer Science *GPA:3.85/4.0*

PES University, Bangalore, India **Aug 2015-May 2019**
Bachelor of Technology, Computer Science and Engineering *GPA:8.08/10*

Presentations

Causal Analysis of Corner Kicks
Technologies Used: Python, Causal Inference Libraries, Machine Learning Algorithms, Matplotlib

- Delivered a talk on 'Causal Analysis of Corner Kicks' at the OPTA FORUM 2024 in London. Proposed using a causal framework to estimate the effect of granular features on shot attempts during corner kicks, providing actionable insights for analysts and coaches to identify optimal matchups for attacking and defensive setups. Blog - harshm16.github.io/causal_analysis.html