Harsh Mishra

(312) 776-4684 | mishraharsh169@gmail.com | linkedin.com/in/harsh-mishra-515624144/ | harshm16.github.io/

SUMMARY

Computer Science master's with 2 years of Industry experience as a Systems and Software Engineer at Hewlett Packard Enterprise. Experienced in automating infrastructure and deployment using CI/CD, Data Engineering, application of Machine learning algorithms and Data Visualization.

EDUCATION

University of Illinois at Chicago (UIC) **Aug 2021-May 2023**
Master of Science, Computer Science *GPA:3.85/4.0*
Coursework: MS Thesis, Cloud Computing, Visual Data Science, Causal Inference, Machine Learning, Natural Language Processing

PES University, Bangalore **Aug 2015-May 2019**
Bachelor of Technology, Computer Science and Engineering *GPA:8.08/10*
Coursework: Data Structures, Big Data, Cloud Computing, Machine Learning, Computer Network Security, Data Analytics

TECHNICAL SKILLS

Proficient: Python, C, R, Ansible, SQL
Familiar: Scala, Java, Go, PowerShell, MATLAB, JavaScript, D3js
Tools/Software: Tableau, Docker, AWS (S3, Lambda, EMR, EBS, SageMaker, Kinesis), Hugging Face, Jira, Latex, Postman

LABORATORY WORK

Research Assistant, Computer Science Department, UIC **Sep 2021-Dec 2022**
Technologies Used: Python, Pytorch, TensorFlow, Docker, Linux

- Trained Score based Generative Models using non-Gaussian noise. The paper was selected as a poster presentation at the MMLS 2023 conference. Preprint available on arxiv.org/pdf/2302.02336.pdf. My Masters's Thesis on the same topic can be found on Indigo-UIC: https://doi.org/10.25417/uic.23661801.v1
- Developed an algorithm to convert categorical labels/features to continuous labels, enabling the use of kernel methods for node classification and other GNN tasks. Experiments on Event Stream and Pose Estimation data available on github.com/harshm16/GNN.
- Co-authored a paper on using optimization-based training methods in distributed Machine Learning settings to overcome byzantine worker problems. Preprint available on arxiv.org/pdf/2302.05865.pdf.

WORK EXPERIENCE

Hewlett Packard Enterprise (HPE) - Bangalore, India **July 2019 – July 2021**
Systems and Software Engineer

- Developed scripts to automate the deployment of MLOps as a Service offering. Used REST APIs to mimic deprecated PowerShell functions in Python and used Ansible for automated deployment.
- Simulated cyber threat patterns using graph databases in Neo4j and wrote SQL queries to detect such patterns.

Hewlett Packard Enterprise (HPE) - Bangalore, India **Jan 2019-July 2019**
Software Developer Intern

- Developed and deployed a DataOps pipeline using open-source applications and wrote bash scripts to automate the CI/CD pipeline.

PROJECTS

ETL Pipeline, UIC **Oct 2021-Dec 2021**
Technologies Used: Scala, D3.js, Kafka, AWS - EBS, Spark

- Built and end to end Log Analysis pipeline. AWS EBS – continuously generate logs, Kafka – real time streaming, Hadoop Spark & MapReduce – data crunching, D3.js – dynamic visualization of results. – Code

Receipt Processing Webservice **June 2023**
Technologies Used: Python, Docker

- Created a web service with two API endpoints: one that processes JSON receipts and generates unique IDs, and another that retrieves points awarded for a given receipt ID. The application was containerized using Docker. – Code

Data Visualization, UIC **2022-Dec 2022**
Technologies Used: D3.js, JS

- Used all the stages of SDLC to come up with different ways to visualize amino acid data. The visualization allows the user to dynamically interact with the 3D structures of the protein and shows visual representations of various other properties & interrelations in the amino acid sequence. – Code

Causal Inference, UIC **2022-Dec 2022**
Technologies Used: Python

- Developed a causal analysis pipeline to identify factors that influenced public sentiments during the COVID-19 pandemic. The experiments provide comparison between various structure finding algorithms used to find causal graphs and then utilizes Bayesian Networks to find the conditional probabilities. – Code