Ansh Mehta

Boston, MA (02119) | 656-223-7047 | mehta.ansh1@northeastern.edu | LinkedIn | GitHub

EDUCATION

Northeastern University, Boston, Massachusetts, United States **September 2023 - May 2025**
Candidate for Master of Science in Computer Science
Coursework: Foundations of Artificial Intelligence, Algorithms and Data Structures, Programming Design Patterns, Cloud Computing

The LNM Institute of Information Technology, Jaipur, Rajasthan, India **August 2017 - May 2021**
Bachelor of Technology in Computer Science
Coursework: Machine Learning, Image Processing, Natural Language Processing, Social Network Analysis, Data Sciences (Applied Statistics)

SKILLS

Programming Languages : Python (PyTorch, PySpark, Pandas, NumPy, Scikit-Learn, Matplot, NLTK), R, JAVA, Javascript, SQL, C++
Databases : Cassandra, ElasticSearch, Redshift, PostgreSQL, MySQL, Clickhouse
Toolkits/Framework : Github Actions, Kafka, Git, Docker, Kubernetes, Ansible, Terraform, Jenkins, Azure Cloud
Data Science/Analysis Techniques: Regression, Data Mining, Clustering, SVM, Deep Learning (RCNNs, LSTM), Reinforcement Learning, LLMs
Certifications: Data Science Foundations, Python for Data Analysis, Managing Data with SQLite

EXPERIENCE AND INTERNSHIPS

Software Engineer (Data Engineer) **July 2021 – July 2023**
Whatfix, Bengaluru, India

- Engineered a large-scale data processing and web analytics platform by integrating JAVA (Spring boot) microservice, Nginx, Clickhouse, Apache Kafka, Azure VMs, streamlining the ingestion and transformation of 50M events data per day
- Orchestrated development of intricate data pipelines with Terraform, successfully integrating data from disparate sources into Clickhouse database, enhancing data processing efficiency
- Designed materialized views within Clickhouse, employing advanced features to optimize 30 distinct data tables. Produced virtual tables for resource optimization and clustering keys in SQL to enhance query performance
- Optimized data science workflows by connecting ClickHouse DB with Apache Superset using Apache Spark, facilitating quicker insights and decision-making to product teams, also wrote DevOps automation scripts for improving efficiency
- Fostered collaboration by query resolution sessions on analytics product for solutions and success teams addressing over 150 specific inquiries related data visualization strategies, data integrations, leading to 40% reduction in support tickets

Software Engineering Intern **January 2021 – June 2021**
Whatfix, Bengaluru, India

- Facilitated an organizational migration of backend services to Azure from AWS, handling weekly interactions from analytics team, that enhanced efficiency and reduced costs by optimizing cloud resources
- Devised and initiated improvements to an ETL pipeline using Hadoop (PIG) to process analytics event data encompassing data cleansing, transformation, and loading the refined data into Amazon Redshift database in an serverless manner
- Created reusable 10+ React components for tabular and chart data aiding users in analyzing critical KPIs of adoption and usage of Whatfix platform offering stakeholders insights on operational efficiency

Research Intern **December 2019 – December 2020**
The LNM Institute of Information Technology, Jaipur, India

- **Publication:** ***Hybrid Computing Scheme for Quasi-Based Deployment in the Internet of Things***
- Proposed clustering (K-means, Hierarchical clustering) using python to evaluate deployment of IoT devices over a large region, and publishing a quasi-random model for deployment and localization of devices
- Researched and applied distributed computing scheme to precisely enable computation and communication in devices, helping develop test framework for evaluating the deployment over multiple network parameters
- Expanded research to supervised learning models (Support Vector Machine, Random Forest), and AI model to enable classification of devices in client-server computing scheme for improving utilization of network resources
- Achieved a 10% improvement over novel model of deployment, and aiding in effective communication of IoT devices through machine learning (AI/ML) techniques

ACADEMIC PROJECTS

Reinforcement Learning in NES Tetris: Deep-Q-Networks, Double DQN, Q-learning **September 2023 – December 2023**

- Implemented reinforcement learning model (Deep-Q-Networks) using python on training Artificial Intelligent agent to play tetris
- Preprocessed and scaled environment inputs using OpenCV, reducing noise for learning optimization

Entity Relationship Extraction in Fictional Novels: Sentiment, Relationship of Entities **February 2019 – April 2019**

- Developed a NLP model capable of handling linguistic features and extracting meaningful insight from medical and patient data
- Utilized Python libraries like NLTK and spaCy, to perform text analytics, tokenization, PoS tagging, and data visualization

Exploring Network of US Flights: Complex Network Analysis **September 2020 - December 2020**

- Prepared visualizations using Python on various network parameters (centrality, distributions, clustering coefficients, PageRank)
- Performed statistical modelling, and visualizations and comprehending the reports to give meaningful insights into the US flights