PRATEESH REDDY

prateeshreddy99@gmail.com || GitHub || LinkedIn

EDUCATION

Indiana University Bloomington, ***GPA: 4/4***
Bloomington, IN, US
Master Of Science in Data Science Jan 2021 – Jan 2023
Course Work: Software Engineering, Advance Database Concepts, Elements of Artificial Intelligence, Deep Learning systems, Data Mining, Applied Algorithms, Big Data Management, Statistics, Machine Learning

Gandhi Institute of Technology and Management University ***GPA: 8.4/10***
Hyderabad, India
Bachelor of Technology in Computer Science and Engineering
Relevant Course Work: C, C++, Java, Python, Software Engineering, Data Structures and Algorithms, Data Mining, Machine Learning, Linear Algebra, Statistics

WORK EXPERIENCE & INTERNSHIPS

Data Science Engineer, CYIENT Limited **Hyderabad, India**
Technologies: Python, R, Java, Apache Spark, Snowflake, ReactJS, Javascript Dec 2019 – Jan 2021

- Extracted road sign boards through an object detection model from terrestrial imagery to minimize manual efforts of data annotation for North American-based clients.
- Updated the model to increase concurrency and fault tolerance and achieved a decent hit rate resulting in a saving of 12 FTEs.
- Developed an application to calculate the dimensions of Poles via 3D image which resulted in huge cost reduction.

Data Scientist Intern, Alexa AI Amazon **Santa Barbara, CA**
Technologies: Pyspark, Python, Java, AWS, Apache Kafka May 2022 – August 2022

- Low-frequency Alexa questions had higher confidence intervals/margin of error. Performed Stratified sampling using various distributions and presented their differences. The result led to a change in Alexa tables' data pipeline.
- Improved Alexa categorizer model by extracting relevant data based on the features using AWS Glue to load them with the existing S3 buckets.
- Created a dashboard for historical comparison of Alexa's successful answering metric with the target and baseline which is now being used in Alexa-wide weekly webinars to discuss the overall performance.

Research Data Analyst, Indiana University **Bloomington, IN, US**
Technologies: Python, R, Tableau, SQL, Microsoft SSIS Jan 2021 – Jan 2023

- Developed data pipelines and Tableau reports for providing insights into student performance, retention, and graduation distribution. Employed data-driven insight to make key decisions on approving pre-requisite course waivers.
- Developed reports on the Indiana University grad school website, Improved efficiency by adding additional data integration and transformation to their data warehouse.

SKILLS

- **Programming:** Java, Python, R, NoSQL, SQL, Pyspark, JavaScript, C++, C, SQL
- **Web:** Flask, Django, React JS, HTML, CSS, REST, Microservices
- **Databases:** MongoDB**,** Snowflake, MySQL, Azure Databricks, Microsoft SSIS
- **Cloud Distributing Services:** Amazon Web Services (AWS), DevOps
- **Tools:** Apache Spark, Tableau, Linux, Git, Jenkins, RabbitMQ, Docker/Kubernetes

ACADEMIC PROJECTS

Cognitive Search Engine by NLP:

- Provided Cognitive search capability for *Eli Lilly and company* to search against databases like FDA and EMA via natural language questions and return relevant results to help with accelerating regulatory submissions for Eli Lilly.
- Performed abstraction-based Natural language generation methods like T5 Transformer, GPT- 2 Algorithm, and BART Transformer.

Face Generative Adversarial Networks (GANs):

- This Master's project solves the issue of misclassification even for the popular models by the ability to generate millions of relevant image data. The uniqueness of this project lies in generating any required style of data using Neural Style Transfers.

PixelGram Application:

- Designed a scalable distributed photo-sharing application for researchers to manage and share photos
- Architected and developed 5 microservices for application with Rest APIs and RabbitMQ as intra-service communication
- Implemented API gateway as middleware to validate/authenticate requests through JWT tokens and to act as a single endpoint
- Deployed dockerized micro-services into OpenShift cluster scripting resource files through Jenkin pipelines on IU Jetstream cloud
- Achieved up to 1800 concurrent requests per instance, tested through load test in JMeter

PUBLICATIONS & LEADERSHIP EXPERIENCE

- Published a research paper on Protecting banking Transactions using 'Blockchain Technology without tokens' in International Journal on Emerging Technologies ISSN No. 0975-8364
- Finished the Course "Grokking Modern System Design Interview for Engineers & Managers"
- Given a GUEST LECTURE on the graduate level course about TABLEAU technology along with NORIKO HARA Z637: Information Visualization on March 2021.