Aman Tej Vidapu

• +1(336) 493 1967 • Raleigh, NC • amantejv7@gmail.com

PROFESSIONAL SUMMARY

Results-driven Data Scientist with nearly 3 years of experience building scalable data pipelines, implementing NLP models, and leveraging LLMs to drive insights for better business decisions. Proficient in cloud computing, distributed systems, and machine learning, with expertise in deep learning, time-series forecasting, and statistical analysis. Proven ability to optimize operations, enhance model performance, and deliver actionable recommendations through data visualization.

EDUCATION

University of North Carolina Greensboro | Master of Science in Computer Science | Concentration: Data science and Big Data. Organizations: Graduate student Association (Senator) Greensboro, NC | Aug 2021 - Dec 2022

Osmania University | Bachelor of Engineering in Computer Science Hyderabad, India | Aug 2017 - June 2021

WORK EXPERIENCE

Data Scientist I Jan 2023 – Oct 2024
LexisNexis | Raleigh, NC

- Designed **scalable** data pipelines using Scala, Spark, and Python, improving data processing for **NLP** solutions and delivering actionable insights to stakeholders for better business decisions.
- Launched a new feature utilizing a transformer-based model for co-reference resolution and information/entity extraction, achieving a **75%** F1 score. This feature significantly improved the product's coverage across diverse datasets, ensuring enhanced performance and broader applicability across multiple jurisdictions.
- Replaced multiple legacy data pipelines with modern models utilizing **AWS** cloud options, processing over 20 million legal documents and reducing processing time by **30%**, resulting in improved efficiency and reduced complexity in data handling.
- Engineered scalable ETL processes to manage **20 million metadata records** using distributed computing and parallel processing, optimizing data lake performance and reducing endpoint load.

Graduate Research Assistant Mar 2022 – Dec 2022
University of North Carolina Greensboro | Greensboro, NC

- Developed bounding boxes to precisely track the center of nuclei in chromatin microscope images using **CNN** and **R-CNN** models, achieving a **4.5%** accuracy improvement in Faster R-CNN through advanced data augmentation techniques, enabling detailed spatial analysis.
- Utilized **LSTM** models to predict the next position of nuclei as a time series problem, enhancing tracking of nucleus movement over time and supporting dynamic analysis of chromatin structure.

PROJECTS

Cyclist: Google Analytics Case Study | R, Tableau, SQL

- Performed **EDA** in **SQL**, **R** to analyze rider patterns, revealing key insights that could increase rider retention by up to 10**%**.
- Developed **Tableau dashboards** to present findings, leading to actionable recommendations for converting casual users into long-term members and potentially boosting subscription rates.

COVID-19 Real-Time Data Analysis | Statistical knowledge, Python, Dash, Git, and Machine learning.

- Cleaned and formatted data, merging with enrichment sources to improve data quality by **15%** and reveal hidden patterns.
- Applied feature selection techniques and hypothesis testing (t-tests, p-tests) to identify key factors. Developed regression models with 87**% accuracy** for predicting COVID-19 trends.
- Created a user-friendly **Dash web app** for real-time data visualization, improving user engagement and facilitating data-driven decision-making.

SKILLS

- **Programming Languages:** Python, SQL, Scala, R, Java
- **Frameworks**: TensorFlow, Keras, PyTorch, Flask, React, FastAPI
- **Tools**: **AWS** (S3, EMR, SageMaker), Docker, Git.
- **Machine Learning**: Classification, regression, clustering, deep learning, time series analysis, transformers, LLM's.
- **Databases**: PostgreSQL, SQL Server, DynamoDB

ACHIEVEMENTS AND CERTIFICATES

- IEEE CS PROGRAMMING LEAGUE 2020: Second place
- Google Data Analytics Professional Certificate