Mihir Harjivan Padsumbiya

14694283574 | padsumbiyamihir@gmail.com | Linkedin | Github

EDUCATION EXPERIENCE

Master of Science, Business Analytics – Data Science **3.96/4.0 GPA**
The University of Texas at Dallas – Awards: *Dean's Excellence Scholar* August 2022 – May 2024
- Certified specializations: Applied Machine Learning, Healthcare analytics

CERTIFICATIONS

Microsoft Azure AI Fundamentals, Snowflake Core, Alteryx Micro-foundational, AWS Training Cloud Practitioner

SKILLS

Languages/Databases: Python, PostgreSQL, R programming, JSON, SAS, SQLite
Tools: MS Excel, Tableau, Microsoft Power BI, Atlassian, Alteryx, Git VC, JIRA, Bitbucket
Machine Learning/Cloud Platforms: Snowflake, Apache Spark, Microsoft Power Platform, Microsoft Azure, Docker, Nextflow pipelines, PyTorch, TensorFlow, ML Ops, NLP, Scikit-Learn, Time Series analysis, ETL/ELT architecture, Econometric modeling, Data Pipelines, End to end Machine learning pipelines

WORK EXPERIENCE

LANTERN PHARMA (NASDAQ: LTRN) Data Scientist - NLP(Jan 2024 - Current)

- Implemented Extract, Transform, and Load (ETL) processes using SQLite and Python. Curated, formatted, and structured big data from disparate sources like web, clinical tests, research articles, aligning with best data engineering practices.
- Generated and initiated novel research questions on the use of NLP for reducing pre-clinical research timelines.
- Built novel Natural Language models using custom neural networks, random forest, and **fine-tuned transformers like PubMed BERT**, in an ensemble setting for innovative oncology drug scrapping projects with an accuracy of 95%.
- Incorporated Azure Co-pilot LLMs into Lantern Pharma's research workflow, automating the extraction and synthesis of scientific texts, which enhanced cancer drug discovery processes and increased research output by 30%.
- Adopted best practices in source version control using GitHub, with appropriate documentation in Agile framework.

ENERGYBOX INC., Nashville, TN Data Analyst Intern (May 2023 – Aug 2023)

- Developed pipelines for automatic data extraction (JSON) and pre-processing with Python algorithms saving 35 hours/wk.
- Applied AWS Textract and NLP techniques for deep learning-based OCR model in Python, saving 40+ hours/week by automating bill extraction flow driving broader quarterly goals for the team.
- Empowered product managers by creating 30+ Python and Power BI dashboards on SaaS data, driving enhanced workflow, stronger stakeholder relationships, and improved ROI for clients through platform data insights.
- Increased ARR by 100% by collaborating across multiple teams including global DA, sales, and customer success.
- Implemented best practices in project tracking using Atlassian, Bitbucket, and Confluence, reducing sprint times by 10%.
- Recognized in 6 out of 13 weeks for exceptional dedication, voluntarily extending after-hours support to teammates across multiple teams, significantly contributing to meeting company deadlines and sprint targets.

THE UNIVERSITY OF TEXAS AT DALLAS, Richardson, TX Teaching Assistant- Statistics (Aug 2023 – Dec 2023)

- Instructed graduate-level courses in Advanced Statistics for data science and Decision-Making, imparting students with statistical acumen in regression analysis, hypothesis testing, and data interpretation, critical skills, using R, and SAS.

HACKATHON / CONSULTING EXPERIENCE

Lantern Pharma Hackathon Nov 2023

- Secured first place in the Lantern Pharma Hackathon amongst 195+ participants by developing a BiLSTM + Attention-based CNN model, ensembled with rule-based NLP techniques, to identify and extract over 8,300 drugs from 100,000+ public research articles with ~80% accuracy, outperforming the runner-up by a significant margin.

Conagra Brands Student Consultant – Data Analyst (Jan 2023 – May 2023)

- Increased revenue potential by 5% by identifying **critical** interactions and opportunities across categories through advanced data engineering and quantitative analysis on sales data sets using MS-SQL, Python, Alteryx, and SAS.
- Developed Machine Learning models to identify consumer behavior, informed decision-making in a team environment.
- Queried 10+ insights using MS-SQL and communicated key performance metrics using causal inference to executives.

PUBLICATIONS / PROJECT EXPERIENCE

Automatic crack detection using Convolutional Neural network (ML) – Survey Lab, NU July 2020-Dec 2021
Journal of Soft computing in Civil Engineering (ISSN: 2588-2872) *DOI :* 10.22115/SCCE.2022.325596.1397

- Developed a Novel CNN model using Deep Learning for Crack detection capable of processing low pixel-density images with an accuracy of 97.8%. Implemented the model using Python libraries Keras and TensorFlow.

Predicting Mental Health amongst teenagers Aug 2023 – Dec 2023

- Spearheaded the development of a production ready Gradient Boosting Trees predictive model for identifying mental health issues in teenagers, with a comprehensive analytical plan for data ingestion, model refinement, and deployment.