

# Mihir Harjivan Padsumbiya

14694283574 | [padsumbiyamihir@gmail.com](mailto:padsumbiyamihir@gmail.com) | [Linkedin](#) | [Github](#)

## SKILLS

**Programming Languages:** Python, SQL, R programming, SAS, DAX  
**Databases and Data Warehouses:** MySQL, PostgreSQL, DynamoDB, Snowflake, RedShift, Databricks, ELT Architecture  
**Data Processing and Collaboration:** Apache Spark, Apache Kafka, Airflow, GitHub, BitBucket, Atlassian JIRA, Agile Practices  
**Data Visualization and BI Frameworks:** MS Power BI, Alteryx, Tableau, QuickSight, ggplot2, Matplotlib, Plotly  
**Machine Learning and Cloud Computing:** Microsoft Azure, Lang Chain, Docker, PyTorch, TensorFlow, Pandas, NumPy, SciPy, Scikit Learn, Streamlit, MLOps, Machine learning models  
**DS Concepts:** Pricing Analytics, A/B Testing, Generative AI, NLP, Time Series forecasting, Econometric modelling

## CERTIFICATIONS & LICENSES

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

## EDUCATION EXPERIENCE

**Master of Science, Business Analytics – Data Science** **3.94/4.0 GPA - Highest Distinction**  
The University of Texas at Dallas – *Awards: Dean's Excellence Scholar* August 2022 – May 2024  
- Certified specializations: Applied Machine Learning, Healthcare analytics

## WORK EXPERIENCE

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

- Empowered product and customer relationship managers by creating 10+ Power BI Dashboards and predictive models, improving ROI for clients by 10% on average, and driving enhanced analysis and communication with stakeholders.
- Developed and managed ETL (Extract, Transform, Load) pipelines using Python algorithms for automatic data collection and cleaning through REST APIs, JSON files, and databases, resulting in significant time savings of 100+ hours/week.
- Increased recurring revenue by 100%, by collaborating with teams across different divisions, locations, and time-zones.
- Earned special recognition in 6 out of 13 weeks for innovative ideas and voluntarily extending after-hours collaborative support to teammates across multiple teams, improving critical thinking, oral and written communication skills.

**LANTERN PHARMA (NASDAQ: LTRN)** **Data Scientist – NLP & Generative AI (January 2024 - Current)**

- Implemented Extract, Transform, and Load (ETL) pipelines using SQLite and Python. Collected, formatted, and cleaned big data from disparate sources including web, SQL, and NoSQL sources, aligning with best data engineering practices.
- Developed fine-tuned LLMs (Open AI GPTs, Claude 3 etc.) for scalable Gen AI solutions, generating 10+ high value datasets and improving prediction accuracy of a proprietary model to 97 from 80%.
- Leveraged Lang chain framework to enhance Gen AI products increasing accuracy by 2% and run-time efficiency by 10%.
- Adopted best practices in version control and collaboration using GitHub, contributing to production code environment.

**THE UNIVERSITY OF TEXAS AT DALLAS, Richardson, TX** **Teaching Assistant- Statistics (August 2023 – December 2023)**

- Instructed graduate-level courses in Advanced Statistics for data science and Decision-Making, imparting students with acumen in statistical methodologies and quantitative modelling, including regression, and hypothesis testing, using R.

## HACKATHON / CONSULTING EXPERIENCE

**Lantern Pharma Hackathon: *GitHub*** **November 2023**

- Secured first place in the Lantern Pharma Hackathon amongst 195+ participants by developing a Bi-LSTM + Attention-based CNN model, ensembled with rule-based NLP techniques, to identify and extract over 8,300 drugs from 100,000.

**Finance Hackathon: Predicting Customer Churn using Machine Learning: *GitHub*** **April 2024**

- Led a team to secure first place among 100 competing teams at the prestigious finance hackathon.
- Engineered an end-to-end machine learning product, using a robust ensemble classification and feature selection model of random forest, and XGBoost algorithms, and achieved sensitivity of 89% outperforming the leading model by 8%.
- Developed a Streamlit app for real-time predictions with LLMs for enhanced consumer insights and user experience.

**Conagra Brands** **Consultant – Data Analyst (January 2023 – May 2023)**

- Increased revenue potential by 5% by identifying critical interactions and optimizing key performance indicators (KPIs) and operational metrics through a predictive model on sales data sets using Python, Alteryx, and MySQL.
- Queried 10+ insights using SAS and communicated key performance metrics using a/b testing to stakeholders.
- Identified 2 new products using predictive analytics for market leading offering.

## PUBLICATIONS / PROJECT EXPERIENCE

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

- Developed a Novel computer vision 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.