PREET MODI

Houston, TX | +1 (812) 318-2011 | Mail | LinkedIn | Portfolio

EDUCATION

Master of Science in Data Science — Aug 2022 - May 2024
Indiana University Bloomington | **GPA: 3.85/4** — Bloomington, Indiana
Relevant Coursework: Data Mining, Applied Algorithms, Software Engineering, Machine Learning

Bachelor of Technology in Information Technology — Aug 2018 - May 2022
Dharmsinh Desai University | **GPA: 4/4** | CSI-President — Gujarat, India

TECHNICAL SKILLS

Languages: Python, SQL, R, HTML, CSS,React JS
Databases: MySQL, SQL Server, MongoDB, PostgreSQL, NoSQL, AWS, GCP, Azure, Oracle, EC2, T-SQL
Data Science : NumPy, Pandas, TensorFlow, Keras, Scikit-learn, LangChain, LangGraph, OpenAI, Databricks, Docker
Statistics: EDA, SAS, Predictive Modeling, Regression, Classification, Time Series Analysis, Hypothesis Testing, MLOps
Other Tools: CI/CD, ETL/ELT pipelines, REST APIs, Dynamics 365, SAP, GitHub, FIS, Enuit, UI/UX, GenAI, GIS

WORK EXPERIENCE

Data Scientist — Sept 2024 - Present
CenterPoint Energy | *ETRM, Python, Riskonnect, BigQuery, Dataverse, ML, AWS* — Houston, TX

- Developed robust **AWS (S3, Redshift, Databricks)** pipelines for processing and analyzing 50,000+ energy portfolio positions, enabling real-time risk management for trading strategies.
- Performed **EDA**, built scalable **ETL pipelines** in Python to analyze and process power trading and safety data.
- Implemented **GenAI models** (BERT, GPT-4) to forecast commodity derivatives (Futures & Options) and portfolio risk.
- Developed **REST APIs** to integrate trading and ETRM platforms, enabling automated data flows via Dataverse.

Research Data Scientist — Aug 2022 - Jul 2024
Indiana University Bloomington | *Power BI, SQL, IBM Cloud, Docker, Data Visualization* — Bloomington, Indiana

- Analyzed data, processed large datasets (>10M records) and created novel visualizations for **Carnegie Classification.**
- Migrated research analytics workflows from AWS Sage Maker to Docker-managed container environments, strengthening MLOps practices and improving pipeline reliability, resulting in 35% faster data processing.

Data Science Intern — May 2023 - Sept 2023
Sacoma Specialty Products | *SQL, Epicor, Redshift, SAP, MySQL, Quicksight, Alteryx* — Edinburgh, Indiana

- Integrated **ERP** systems with **AWS**, designed databases, and implemented BAQs to ensure seamless data migration.
- Executed queries in **SQL** and deployed a **Random Forest predictive model** to forecast manufacturing orders in ERP.

PROJECTS

GenAI-Powered Commodity-VaR Insights App | *Python, Pandas, Airflow, GPT-4, ETRM, OpenAI, API Development*

- Processed energy trading & portfolio data with **Python**, to compute returns, volatilities & correlations for risk analysis.
- Built **ETL** workflows with **Airflow** to transform ETRM data for VaR and scenario-based risk assessment for hedging.
- Engineered **Parametric, Historical, and Monte Carlo VaR** models, cutting compute time 30% via parallel processing.
- Integrated **GPT-4** by writing Python scripts to call the **OpenAI API**, generating real-time risk summaries and scenario insights for portfolio analysis in **Dash, Plotly** and **Tableau.**

High Energy Incident Forecasting and Safety Power App | *Python, Power Apps, Dataverse, BigQuery, Scikit-learn, LLM*

- Engineered a Python pipeline to predict high-risk energy incidents using weather and safety data; trained **Random Forest** and **Logistic Regression**, achieving 82% F1-score.
- Deployed model outputs via **Azure** and integrated predictions into a Power Apps-based safety app, leveraging **Dataverse** and **BigQuery** for real-time reporting and analytics.
- Developing an **LLM** chatbot with **Google Vertex AI** to assist CenterPoint field employees with high-energy queries.

PUBLICATIONS

Insurance Management with Premium Prediction | *DOI*

- International Journal for Research in Applied Science and Engineering Technology (IJRASET)(Impact Factor: 7.429)

An efficient Artificial Neural Network for Coronary Heart Disease Prediction | *DOI*

- International Journal for Research in Applied Science and Engineering Technology (IJRASET)(Impact Factor: 7.429)