Harshkumar Modi

modi.hars@northeastern.edu | http://www.linkedin.com/in/hkmodi31 | +1 (267) 356-7063 | GitHub

Data Science MSc candidate at Northeastern University, with extensive experience in data engineering, data analysis, machine learning and cloud computing. Actively seeking a full-time position that demands skills in Python, SQL, AWS, and Tableau along with the curiosity to learn and the skill to quickly adapt to new technologies.

TECHNICAL SKILLS

Analytical: Predictive Modeling, Natural Language Processing (LLMs), Time Series, Data Analysis, Warehousing, Computer Vision
Programming: Python (numpy, pandas, matplotlib, seaborn, scikit-learn, TensorFlow, pytorch, Keras, nltk, spacy), R, SQL (MySQL, PostgreSQL), NoSQL (MongoDB), A/B Testing, Docker.
Cloud computing: AWS Suite (IAM, S3, EC2, Glue, Lambda, QuickSight), GCP, Apache Spark (PySpark).
Software: Git/GitHub (Version Control), Airflow (Workflow orchestration), Tableau (Data Visualization), Figma (UI/UX design)

EDUCATION

Northeastern University, Boston, MA — Jan 2022 – May 2024
Master of Science in Data Science — GPA 3.95
Relevant coursework: Data Management and Processing, Supervised Machine Learning, Algorithms, Cloud Computing, Unsupervised Machine Learning, Database Management Systems, Natural Language Processing, Large Language Models.
Teaching Assistant: Discrete Mathematics (CS 1800), Natural Language Processing (CS 4120 / 6120), Database Design (CS 3200).

SRM Institute of Science and Technology — Jul 2017 – May 2021
Bachelor of Technology in Computer Science Engineering — CGPA 8.38
Relevant coursework: Advanced Calculus and Complex Analysis, Statistics and Probability, Data Structures, Data Science and Big Data Analysis, Software Engineering Principles, Object Oriented Analysis and Design, Computer Networks.

WORK EXPERIENCE

COI Energy, Boston, MA — May 2023 – Aug 2023
Energy Analyst - Intern
- Conducted in-depth analysis of energy and battery policies, driving strategic growth across 3 new state markets.
- Refined app UI with Figma, improving usability for Google Play Store release, positively impacting user engagement.
- Utilized ARIMA and Prophet to perform time series analysis to predict energy usage of 20+ clients with an accuracy of 91%

Salient Predictions, Boston, MA — Jan 2023 – Apr 2023
Machine Learning Engineer - Intern
- Enhanced climatology trend analysis and anomaly detection by manipulating multi-dimensional data with Python and xarray.
- Boosted medium-long range forecast reliability by engineering accuracy plots from refined data, elevating prediction precision.
- Achieved a 60% increase in web app visualization efficiency by integrating a python API with user interface components.

PROJECTS

Optimization of Conversational Summarization Using BART Models (*GitHub*) **- Python, Transformers, Hugging Face**
- Evaluated three text summarization models, achieving optimal results with fine-tuned BART model on SAMsum dataset.
- Conducted fine-tuning of Facebook's BART model, enhancing performance significantly on custom conversation summarization.
- Utilized Rouge evaluation to benchmark model performance, outperforming pre-trained model by 2x using AWS Sage Maker.

Grocery Store Sales Analysis using PySpark on Databricks (*GitHub*) **– Databricks, PySpark**
- Utilized PySpark to conduct an in-depth analysis of 10,000 rows of retail sales, uncovering critical purchasing patterns and trends.
- Mapped out regional sales trends, enabling the strategic optimization of inventory distribution and targeted marketing efforts.
- Employed Apriori algorithm to discern trends in customer orders, refining product recommendation strategies.

Boston Airbnb Database Management and Visualization (*GitHub*) **- Python, PostgreSQL, Tableau**
- Developed a python and API-driven ETL pipeline for updated Airbnb data, streamlining analytics for Boston's Airbnb trends.
- Executed PostgreSQL scripts to ensure 100% data integrity and efficient storage in designated tables for Airbnb analytics.
- Utilized Tableau to create detailed data visualizations, providing comprehensive reports of the Boston Airbnb market.

Aviary Analytics: Cloud-Based Bird Strike Data Management and Analysis (*GitHub*) **– R, MySQL**
- Accomplished a cloud-based relational DB, using accurate AWS RDS configuration, by integrating R and SQL for data analysis
- Enabled advanced data analytics on bird strikes, as evidenced by insightful reports, through R programming and SQL querying.
- Streamlined data ingestion from CSV to SQL, indicated by a 100% data accuracy, using R scripts for database population.

YouTube Data Analysis – Python, AWS (Glue, Lambda, S3, QuickSight)

- Engineered an AWS Lambda ETL pipeline, converting S3 data lake content to parquet, increasing data retrieval efficiency.
- Automated data refinement with AWS Glue crawlers/triggers; streamlined data storage management after lake updates.
- Synthesized AWS QuickSight dashboards to illustrate data trends, bolstering analytics with insightful visualization.

Financial market analysis using supervised and unsupervised machine learning - Python

- Analyzed financial markets with PCA, Lasso, and KBest, isolating performance indicators across varied stock categories.
- Composed statistical quantitative functions to enrich raw dataset with new features set for advanced market analysis.
- Validated LSTM model predictions on multiple stable and unstable stocks, concluding varied results of using different algorithms.

Real Time License Plate Recognition using Convolution Neural Networks (*GitHub*) - **Python**

- Curated data to limit and identify license plate boundaries from overall image of car and noise in data using HaarCascade.
- Built a CNN network to train network on previous license plate data to capture the Indian license plate characters.
- Applied model on live video stream to extract and determine characters and run it by registration API to get details of vehicle.

RESEARCH EXPERIENCE

Melanoma Classification using Convolutional Neural Network (CNNs) Integrated with Tabular Model - (*publication*)

- Implemented machine learning techniques to enhance wireless capsule endoscopy analysis, improving diagnostic efficiency.
- Utilized TensorFlow and Keras for model development, contributing to improvements in gastrointestinal disease detection.
- Developed a CNN model achieving 97.82% accuracy in classifying 13 digestive tract abnormalities using endoscopy data.

Digestive Tract Abnormalities Classification Using Endoscopy Data on Convolution Neural Network (CNNs) - (*publication*)

- Conducted an extensive survey of melanoma classification techniques, meticulously assessing their strengths and weaknesses.
- Engineered an algorithm harnessing 'DenseNet 201' and transfer learning, overcoming limitations affected by limited datasets.
- Led multi-layer model development, achieving 96% accuracy in melanoma classification through medical image analysis.