

Harsh Makwana

+1 (980) 327 9431 | [LinkedIn](#) | [GitHub](#) | [Email](#) | Long Beach, CA

SKILLS

Languages: Python, MySQL, JavaScript, R, C++, SAS, VBA, MATLAB
Frameworks: Sklearn, PyTorch, TensorFlow, Keras, Pandas, NumPy, OpenCV, SDK, CLI, Seaborn, Matplotlib, Plotly Dash
Tools & Services: AWS, Microsoft Azure, GitHub, Power BI, Tableau, CI/CD, Spark, Docker, Kubernetes, Heroku, SSIS/SSRS, Alteryx
Applications: Data Science, Machine learning, Data Analysis, Cloud Computing, Deep Learning

EXPERIENCE

Freelance Data Scientist | Belivonce Team - Remote

Jun 2022 - Present

- Executed massive web scraping with Python (BeautifulSoup, Selenium, Requests), successfully compiling over 50+ databases.
- Carried out data processing using Alteryx/Python, coupled with automation in VBA (massive data crunch and directory tools), enhancing data efficiency by 25-30%.
- Developed 10+ ML models focusing on regression and classification within AWS SageMaker using Pytorch, Tensorflow, and Keras.
- Utilized techniques such as over/under-sampling, transformation/pipelines, feature engineering, and ensemble methods, achieving a 16-20% performance increase across various models.
- Achieved a 4.5/5 customer satisfaction rating by presenting analytics with data visualizations/dashboards built via Web-Interface (Plotly Dash & Heroku) and Power BI (DAX).

Data Analyst | Numerator - Vadodara, India

Oct 2020 - Jul 2021

- Revamped data preprocessing time by 40% for major retailers using optimized MapReduce jobs on AWS EMR.
- Developed an XGBoost customer Churn model, boosting customer retention by 15% and lifetime value by 10%.
- Designed Tableau dashboards for monitoring customer demographics, behaviors, and segmentation, enriching strategic business initiatives.
- Prepared technical specifications and documentation about workflow in AWS, aiding future projects.

Machine Learning Engineer | Department of Computer Engineering - SVIT, India

Aug 2019 - May 2020

- Optimized data pipeline for Large-scale consumer dataset utilizing Apache Spark Cluster, reducing processing time by 30%.
- Developed a predictive model for time series analysis, leading to a 10% sales increase.
- Implemented various models (LSTM, SARIMA, ARIMA), resulting in a 6% improvement in accuracy.
- Deployed the model using Databricks with API configuration and testing, resulting in an 8% performance boost.
- Streamlined deployment with Docker and Kubernetes, achieving a 25% increased reliability via Azure CI/CD pipeline.
- Collaboratively integrated ML insights with Power BI dashboards, enhancing stakeholder engagement by 20%.

Data Analyst | Newport Land Enterprise - Vadodara, India

Jan 2018 - Dec 2018

- Migrated 1TB ERP data to Microsoft Dynamics using SSIS/DMA, achieving a 99.5% accuracy.
- Streamlined query processing using SQL by strategic indexing and join optimization, slashing data retrieval time by 30%.
- Facilitated analytical support to CXOs through 50+ SSRS reports, yielding \$100K+ annual sales.
- Elevated KPIs by 10% with critical insights from SAS advanced analytics (logistic regression, clustering).

EDUCATION

Masters in Data Science, University of Massachusetts – Dartmouth

Sep 2021 - Dec 2023

Courses: Advance Mathematical Statistics, Advance Database Design, high-performance computation, Data Visualization, Numerical optimization, Business intelligence & knowledge management, statistical analysis, business analytics, Advance Machine Learning

Bachelor of Engineering in Computer Engineering, Gujarat Technological University

Aug 2016 - Aug 2020

Courses: Data Structures, Design of algorithm, Computer Architecture, Computer Networks, Operating Systems, Artificial Intelligence, Big Data Analytics

PROJECTS

Loan Approval Prediction

- Predicted loan approval status, initially achieving 76% accuracy by selecting features through a correlation matrix, and after feature augmentation, the model reached 81% accuracy. **Tech:** Amazon SageMaker, Amazon S3, Seaborn **Lang:** Python

Time Series Prediction Using LSTM

- Forecasted the next 1000 values for an input of N values of a Sin wave using LSTM and after optimizing with limited memory BFGS loss was 0.002. **Tech:** Pytorch **Lang:** Python

Air Quality Index Interactive Visualization

- Created an Interactive web Visualization for Pre and post-COVID AQI of the USA using D3.js and Preprocessed the Topo.json data files using NumPy. **Tech:** MySQL **Lang:** python, JavaScript

Embedded System Design

- Designed a mobile piano with Raspberry Pi, using image recognition of fingers on LED-lit nodes to produce synthesized audio through a speaker. **Tech:** OpenCV **Lang:** Python

SQL Query Engine

- Developed a SQL query engine using JSQL parser, JDBC, Evallib, and lang3 libraries. Delivered seamless execution of SQL commands and nested queries.
Tech: MySQL **Lang:** JavaScript

Predicted the Sale of Automobiles to Target Customers

- Build a predictive model to classify car sales with an 89% accuracy and visualize performance by a confusion matrix.
Tech: SciKit-Learn, Pandas, NumPy, Matplotlib **Lang:** Python