# Sparsh Marwah

Boston, MA 02130 | marwah.sp@northeastern.edu | +1 (857) 225-9142 | LinkedIn | GitHub | Portfolio

## Work Experience

**Data Science Analyst**, Tredence Analytics Solutions Pvt. Ltd., Bengaluru, India **Jun 2021 - Jul 2023**

- Executed **A/B and multivariate testing** methodologies, generating insights that informed strategic decisions for e-commerce data, enhancing conversion rates by 15% through targeted consumer behavior analysis
- Developed predictive models using **Python** and **SQL**, leveraging machine learning techniques (Linear Regression, XGBoost, Clustering) to improve customer segmentation, increasing targeting accuracy by 25% and boosting customer engagement by 15%
- Engineered and deployed end-to-end **MLOps** pipelines for automated model retraining, monitoring, and version control using **MLflow** and **CI/CD** workflows, reducing model deployment time by 40% and improving model performance tracking by 30%
- Led data preparation, feature engineering, and exploratory data analysis to support model performance optimization, achieving a 91.2% accuracy
- Crafted interactive business intelligence dashboards in **Tableau**, delivering actionable financial insights that informed pricing strategies and boosted conversion rates by 15%, enhancing overall revenue generation for the organization
- Utilized unstructured and structured data sources to develop data-driven insights for strategic business applications

**Data Analyst Intern**, SJVN Ltd., Shimla, India **Jun 2019 - Aug 2019**

- Collected and analyzed energy inventory data by developing complex **SQL** queries, optimizing stock level assessments and identifying sales trends, leading to a 20% reduction in stockouts and a 15% improvement in inventory turnover
- Drafted data integration workflows documentation to decide the entire lifecycle of the project, ensuring seamless dataflow
- Conducted data quality audits and troubleshooting, ensuring accuracy, integrity, and consistency across datasets, reducing data errors by 30% and enhancing decision-making efficiency by 25%

## Technical Skills

**Programming & Databases:** Python (TensorFlow, PyTorch, Scikit-Learn), SQL (PostgreSQL, MySQL, Hive), MongoDB, Oracle
**Machine Learning & AI:** Supervised & Unsupervised Learning, Predictive Analytics, NLP, Time Series Analysis, Generative AI
**Big Data & Cloud:** AWS (S3, SageMaker, Lambda), Google Cloud Platform (BigQuery, AI Tools)
**MLOps & Deployment:** MLflow, CI/CD Pipelines, Docker, Feature Engineering, Model Industrialization, REST APIs
**Data Visualization & BI:** Tableau, Power BI, Matplotlib, Seaborn
**Collaboration & Reporting:** Git, Jira, Confluence, Microsoft Office Suite
**Certifications:** Python (Programming, Data Structures), Data Science & AI, Intro to Cloud Data Analytics, ETL in Python and SQL

## Academic Projects

**Healthcare Risk Assessment Model** **Jan 2024 - Present**

- Developed a machine learning-based risk assessment model for healthcare underwriting, leveraging structured and unstructured data to optimize risk predictions
- Integrated predictive modeling with clinical review processes, improving underwriting accuracy & reducing manual review efforts
- Implemented NLP techniques for extracting insights from medical notes and text data to enhance predictive capabilities

**Air Quality Prediction** (View Project) **Sep 2024 – Dec 2024**

- Designed and deployed a cloud-native **MLOps** pipeline on **GCP & Airflow** for real-time time-series data ingestion and preprocessing, ensuring efficient data handling and scalability
- Optimized model deployment by integrating **GitHub Actions** & MLflow for automated retraining, reducing deployment time by 50% and maintaining consistent accuracy
- Implemented **REST API** integration to provide real-time air quality predictions, enhancing end-user accessibility and usability
- Automated model monitoring using **MLflow** and **Cloud Functions** for drift detection, ensuring proactive retraining and sustained model performance

**Face Mask Detection** (View Project) **Feb 2024 – Apr 2024**

- Implemented a real-time surveillance system for face mask detection using deep learning techniques, ensuring high-speed, accurate detection.
- Employed semantic segmentation for pixel-level precision and optimized the system for real-time video processing, improving public safety monitoring.

## Education

**Northeastern University**, Boston, MA **Sep 2023 - May 2025**
Master of Science in Data Analytics Engineering, GPA: 3.75/4.0
Relevant Coursework: Data Management in Analytics, Data Mining in Engineering, Machine Learning Operations, Applied Gen-AI
Relevant Experience: Teaching Assistant for **Computational and Visualization for Analytics**

**SRM Institute of Science and Technology**, Chennai, India **Jul 2017 - May 2021**
Bachelor of Technology in Computer Science Engineering
Relevant Coursework: Data Structures, Data Science and Big Data Analysis, Object Oriented Analysis and Design
Publication: AI Music Generator (Research paper)