

HARSH SHAH

119 Sutherland Rd, Brighton, MA 02135 | (908) 636-1325 | shahharsh0612@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Master's candidate in Data Science with 3+ years of experience in machine learning, data analytics, and big data engineering. Skilled in building end-to-end solutions using Python, SQL, and cloud technologies to drive business insights and develop scalable applications.

SKILLS

- **Programming:** Python, SQL, Java, C++, JavaScript, HTML, CSS
- **Machine Learning & Analytics:** Scikit-learn, TensorFlow, ML Algorithms, Predictive Modeling, Time Series, NLP
- **Big Data & Cloud:** Azure (ML Studio, Databricks, Synapse, Blob Storage, Event Hubs), AWS, Spark, Hadoop
- **Data Engineering:** ETL Pipelines, API Integration, Distributed Processing, Stream/Batch Data
- **Data Visualization:** Tableau, Power BI, Seaborn, Matplotlib, GeoPandas, Plotly, Looker Studio
- **Workflow & Tools:** Git, Docker, Jupyter, CI/CD

WORK EXPERIENCE

Petpooja (Restaurant POS Software)

Jun 2025 – July 2025

Data Science Intern

Ahmedabad, India

- Enhanced dashboards for multi-dimensional analytics (sales, inventory, taxation, staff), driving faster decision-making.
- Developed logic for outlet and city-wise data aggregation, improving monitoring efficiency for 100+ outlets.
- Designed models to analyze user behavior and order flows, leading to a 15% improvement in UI/UX efficiency.
- Explored AI models for automated tax categorization, reducing tax reporting errors by 25%.

General Motors

Jul 2022 – Jun 2024

Software Developer

Warren, MI

- Engineered 15+ APIs and automated data pipelines, securing data for 30k users with Grafana-based monitoring.
- Built CI/CD pipelines in Azure to automate builds and deployments, boosting scalability with 85% success rate.
- Applied distributed processing and cloud-native practices to enhance reliability and accelerate development.
- Deployed role-based access controls and input validation techniques to mitigate security threats and enhance trust.
- Delivered seamless remote vehicle control integration, resolving 5–10 software issues daily using Git-based workflows.

General Motors

Jun 2021 – Jun 2022

Senior Quality Assurance Specialist

Warren, MI

- Tested and debugged 400+ APIs (Insomnia, Postman, ReadyAPI), ensuring 90%+ API success during production rollouts.
- Developed 100+ automated/manual test plans in JIRA, improving API reliability and performance validation.
- Trained and mentored junior members in software testing which contributed to a 25% boost in overall team productivity.

Silicon IT Hub Pvt Ltd.

Sep 2020 – Jan 2021

Android Application Developer Intern

Ahmedabad, India

- Built an Android app for truckers to locate nearby parking (10-mile radius), enhancing efficiency and minimizing downtime.
- Tested 5 Android applications based on specific use cases, ensuring streamlined app implementation and enhancing user experience.

PROJECTS

Financial Analytics Web App | [Link](#)

Mar 2025 – May 2025

- Streamlit-based web app that uses sentiment analysis on social media and financial news to assess stock market trends, providing users with stock recommendations based on sentiment correlation.

Predicting Respiratory Hospitalizations from Air Pollution | [Link](#)

Mar 2025 – May 2025

- Built a machine learning model using Azure ML Studio to predict hospitalization risks based on real-time air quality data.
- Designed ETL pipelines using Azure Data Factory & Event Hubs to ingest historical & real-time pollution data.
- Performed time-lag correlation analysis using Distributed Lag Models (DLMS) & Generalized Additive Models (GAMs).
- Created a Power BI dashboard for visualizing high-risk zones & sending alerts via Azure Logic Apps.

Cricket Analytics Web Platform – End-to-End DS & ML Integration | [Link](#)

Feb 2025 – Apr 2025

- Cleaned, structured, and analyzed player-level cricket data using Pandas and Seaborn for impactful performance visualizations.
- Built a KMeans-based clustering model to group players by performance metrics and identify similar profiles using cosine similarity.
- Developed a Flask web app to deliver ESPN-style visual insights, interactive dashboards, and ML-powered player comparisons.
- Integrated role-based insights and top-performer detection across formats, with modular codebase and version control via GitHub.

Exploratory Data Analysis of Boston's District 7's Economic and Population Development | [Link](#)

Oct 2024 – Dec 2024

- Developed a geospatial model for District 7 using census tract shapefiles and data apportionment.
- Empowered the Policy Director with interactive visualizations to assess decade-long socioeconomic and displacement trends.
- Built a live Python dashboard (GeoPandas, Seaborn, Tableau) comparing District 7 metrics to city averages for key indicators.

EDUCATION

Boston University | Master of Science in Data Science | GPA: 3.64

Dec 2025

- Coursework: Machine learning, Data Visualization & EDA, Big Data Engineering, Financial Analytics, AI Ethics, Project Management

Fairleigh Dickinson University | Bachelor of Science in Computer Science | GPA: 3.69

Aug 2017 – May 2021