

## Harnish Shah

Data Scientist/Business Intelligence Analyst/ Data Analyst  
732)351-3241 | [harnishshah25@gmail.com](mailto:harnishshah25@gmail.com) | [LinkedIn](#) | [Github](#)

### **SUMMARY**

Aspiring data scientist with a background in Medical Devices Engineering. As a biomedical engineer, I design, test, and improve innovative and affordable medical devices that can enhance the health and well-being of people and communities worldwide. I have over nine years of experience in the medical devices industry. This has taught me the joys that come with learning, investigation, and collaboration. I've spent the past year as a data scientist and found that I love jumping into problems to help drive businesses decisions or find insight in what may look like messy unstructured data. I have both strong analytical reasoning and communication skills. I look for opportunities to take the initiative and anticipate/resolve problems before they develop. Having recently changed careers, I'm open to work, and excited to dive into business problems to make every day run smoother.

### **SKILLS**

Technologies: Python | SQL | Tableau | Scala | Spark | HTML | AWS EC2 | Jupyter Lab | Github | Excel | Word | PowerPoint | Epic Suite

Skills: Web Scraping | Data Cleaning | Supervised and Unsupervised Machine Learning (ML) | Predictive Modeling | Data Visualization | Data Analysis | Natural Language Processing | Big Query |

Python libraries: Pandas | NumPy | SciPy | Matplotlib | Seaborn | Scikit-learn | Natural Language Toolkit | Beautiful Soup | Praw | Keras | TensorFlow.

Spoken Languages: Hindi (Advanced - Read/Write/Speak), Gujarati (Advanced - Read/Write/Speak), Marathi (Advanced - Read/Write/Speak), Spanish (Beginner – Speak)

### **DATA SCIENCE PROJECTS**

#### **Classification model with NLP to Sentiments of two subreddits**

- Created an unstructured dataset by scraping over 12,000 posts from both subreddits.
- Transformed data in preparation for modeling using feature engineering and NLP.
- Identified topics that were strongly associated with subreddits using a logistic regression classifier.
- Achieved a recall score of .6944 to address the class imbalance in the dataset.

#### **Spotify Data**

- DISPLAY GROUP PROJECT [SPOTIFY INSIGHTS]
- -----
- -----
- -----
- -----
- -----
- -----
- To be added

#### **Regression model to predict home sale prices in Ames, IA**

- Explored a Kaggle dataset of home sales from 2006 to 2010 provided by the Ames's Assessor's office.
- Used data visualization and feature engineering techniques to analyze and manipulate data into a format usable for modeling.
- Evaluated several models using linear regression and LASSO or Ridge regularization.
- Developed a model that was able to predict housing prices with an  $R^2$  score of .97.
- Stood 1<sup>st</sup> of all model submissions on a Kaggle competition. [Kaggle link](#)

## **PROFESSIONAL EXPERIENCE**

### **Soma Technology Intl**

May 2018 – Jan 2023

*Biomedical Engineer, Bloomfield, CT*

- Analyze large amounts of data and develop statistical models to find patterns and solve problems that will help drive strategic business decisions.
- Design and implementation of machine learning solutions using regression, model, clustering (KNN, K-means, Bayesian, Mean-shift), statistical profiling, inference, classification, and predictive analysis.
- Working knowledge of medical device and combination product regulations and standards such as 21 CFR Part 4, 21 CFR 820, ISO 13485, and ISO 14971
- Experience with change controls, deviations, CAPA Experience with Medical Device or Combination products.
- Understanding of the six-sigma process, applying statistical tools to solve statistical problems.

## **EDUCATION**

### **General Assembly, USA – Remote**

Jan 2024 - April 2024

*Certificate of Completion, Data Science Immersive*

### **Saurashtra University, Gujarat, India**

April 2007 – April 2012

*Bachelor of Engineering, Biomedical and Instrumentation Engineering*

## **CERTIFICATES**

### **Databricks**

*Certificate of Completion, [Databricks Lakehouse Fundamentals](#)*

### **Databricks**

*Certificate of Completion, [Generative AI Fundamentals](#)*

### **Databricks**

*Certificate of Completion, Generative AI Engineer (in Process)*

[https://customer-academy.databricks.com/lms/index.php?r=coursepath/deeplink&id\\_path=315&hash=4b1e155390e39095ab2cf0d105df76ec71214dc1&generated\\_by=607554](https://customer-academy.databricks.com/lms/index.php?r=coursepath/deeplink&id_path=315&hash=4b1e155390e39095ab2cf0d105df76ec71214dc1&generated_by=607554)