Harnish Shah

Data Scientist / Data Analyst / ML Engineer
732-351-3241 | harnishshah25@gmail.com | LinkedIn | Github | Portfolio

SUMMARY

I am an aspiring data geek with a background in the Medical Devices Industry. For nearly a decade I worked to design, test and optimize medical devices for communities worldwide. I've sifted through unstructured data to drive business decisions with insights – and I love it! With my analytical reasoning, machine learning and communication skills, I am looking for opportunities to boost efficiency and revenue.

SKILLS

Languages: Python | SQL | R | Scala | Spark
Technologies: Github | Databricks | Tableau | Google Suite | VS Code | Command Line | AWS EC2 | Flask | Jupyter Lab | Docker
Skills: Web Scraping | Data Cleaning | Linear Models | Tree Models | Support Vector Machines | Predictive Modeling | Time Series / Seasonality Analysis | Neural Nets (DNN, RNN, CNN) | Natural Language Processing (NLP) | Principal Component Analysis (PCA) | Pipelines | Data Visualization | Data Analysis | Data Management
Python libraries: Pandas | NumPy | Matplotlib | Seaborn | Scikit-learn | Natural Language Toolkit (NLTK) | Beautiful Soup | TensorFlow / Keras | PySpark

PROFESSIONAL EXPERIENCE

Soma Technology Intl May 2018 – Jan 2024
Biomedical Engineer – Data Analyst, Bloomfield, CT

- Analyze large amounts of down the service costs data and develop statistical models to find patterns and solve problems that will help drive strategic business decisions, which helped bring of the company over 1M annually
- Design and implementation of machine learning solutions using regression, model, clustering (KNN, K-means, Bayesian), statistical profiling, inference, classification, and predictive analysis and it increased the efficiency of field service engineers and the service team by reducing the overall breakdown visits to the healthcare organizations by more than 50%
- 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

NexGen BioEquips Nov 2015 – Feb 2018
Founding Partner, Vadodara, Gujarat India

- Created and executed the marketing plan for penetrating Tier 3 clinics and hospitals in Gujarat
- Undertake projects of hospitals including procurement of medical devices, designing, and setting up various departments such as ICU, OR and ER

VINS Hospital Oct 2014 – Nov 2015
Biomedical Engineer, Vadodara, Gujarat India

- Categorize and schedule device maintenance routine based on risk level to increase patient care while decreasing device downtime
- Negotiate quotes and contracts with external suppliers, which helps to reduce department costs by 30 – 50%
- Implement, and make policies for medical equipment management, such as acquisition, repairs, maintenance, and inventory

GE Healthcare Sept 2013 – Oct 2014
Sales Engineer, Gujarat India

Bankers Heart Institute July 2012– Aug 2013
Biomedical Engineer, Vadodara, Gujarat India

DATA SCIENCE PROJECTS

Recommender System, Traditional Retail Company
Jupyter Lab, Tableau

- Audited, validated, and scrubbed 10M+ rows of sales data and performed the RFM analysis
- Utilized supervised models to get the accuracy of over 97% taking users and the products they purchased
- Built a Recommender System based on user and the products purchased
- Performed seasonality analysis to predict the upcoming sales using the system prepared

Classification model with NLP to Sentiments of two subreddits
Reddit PRAW API, Jupyter Lab

- 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

Hotdog or Not Hotdog
Jupyter lab, Streamlit

- Built and trained a convolutional neural network (CNN) on ~4,200 image files in 6 hours
- Co-created a Streamlit app that accurately classifies uploaded pictures as either hotdogs or not.

Successful Song Prediction with Features
Jupyter Lab, Github

- Cleaned and analyzed 250,000+ rows of Spotify data and built Random Forest, DNN, Logistic Regression, and Decision Tree Classifier models
- Trained a dense neural net model with 79% accuracy in predicting popular Spotify songs
- Collaborated with 3 data scientists over Github for a 20-minute stakeholder presentation

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 R2 score of 0.97
- Stood 1st of all model submissions on a Kaggle competition

EDUCATION

General Assembly, NYC, 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 Fundamentals

Databricks
Certificate of Completion, Databricks Lakehouse Fundamentals

Databricks
Certificate of Completion, Generative AI Fundamentals

WORK AUTHORIZATION

US Citizen (Naturalized)