# Maroti Raghunath Shelke

✉ marotishelke2000@gmail.com | 📱 +91 9075325154
🐙 https://github.com/Marotishelke | 🔗 https://www.linkedin.com/in/maroti-shelke

## SUMMARY

Data Scientist with 3.7 years of experience delivering ML and GenAI solutions across healthcare and industrial domains. Proven expertise in predictive maintenance, and real-time analytics to drive data-driven decisions and operational gains.

## SKILLS

**Programming Languages:** Python, R, SQL, C

**Data Science & Machine Learning:** Regression (Linear, Logistic), Tree-based Models (Random Forest, Boosting),Clustering, Time Series Forecasting (ARIMA, Prophet), Feature Engineering, Model Evaluation, Large Language Models(LLM), Generative AI, Transformers, Word Embeddings (Word2Vec, FastText), Prompt Engineering

**Technologies & Tools:** Pandas, NumPy, Scikit-learn, PySpark, Docker, Git, Azure DevOps, Azure Synapse Pipelines, Microsoft SQL Server, Django.

## WORK EXPERIENCE

**WebMD Internet Brands, Remote**                                                    Jan 2025 - Present
*Data Scientist*

- Built and deployed a production GenAI-powered analytics platform where users can either execute KPI-driven workflows via a Streamlit UI (upload inputs, select KPIs, generate and download results) or interact with a chatbot to receive guided input selection and trigger the same validated analytics pipelines. The platform integrates data from Forian and Vertica databases to support large-scale diagnosis, prescription, and campaign analytics. Implemented RAG using LangChain to ground the chatbot on internal KPI definitions, measurement methodology, workflow rules, and run-specific outputs, enabling accurate execution guidance, post-run explanations, summaries, and follow-up analysis. Automated 18+ analytics workflows, reduced analysis turnaround time by ~80%, and achieved ~99% accuracy compared to legacy analyst reports.
- Applied clinical NLP techniques (NER, text classification, entity linking) on unstructured medical records (patient visits, diagnoses, prescriptions), extracting actionable insights that improved HCP targeting and led to a 20% increase in marketing ROI.
- Designed and executed A/B testing frameworks using Propensity Score Matching and ANCOVA, ensuring unbiased campaign evaluation and measurement.
- Built geospatial clustering models with unsupervised learning (K-means, DBSCAN) to segment HCPs by practice and specialty, improving targeting accuracy and campaign personalization.
- Worked closely with different teams to embed AI insights into client dashboards, helping them make better decisions in pharma marketing.

**Data Scientist – Capgemini**                                                        Aug 2022 - Jan 2025

- Developed a predictive model for silica content in bauxite, using UMAP, Agglomerative Hierarchical Clustering, and PLSR, resulting in 15% better predictive accuracy for ore quality analysis.
- Designed machine learning models for truck refueling prediction, reducing inefficiencies and saving 15 hours per week by accurately estimating remaining fuel and optimizing tank capacity.
- Forecasted mining asset health, predicting machine failures using Predictive Modeling and Anomaly Detection, reducing unplanned shutdowns by 25% and minimizing production losses.
- Created a time forecasting model with Prophet, predicting dust collector health with 94% accuracy, extending maintenance windows and reducing shutdowns from 6 to 2 per year.
- Implemented Quadratic Regression to detect impeller wear and forecast Risk of Failure, increasing component lifespan by 1+ year compared to traditional maintenance practices.
- Developed a Django-based web application for EDA and data visualization, streamlining data analysis and enabling users to gain insights from datasets efficiently.
- Created a code quality assessment tool using Pylint and Rlint for Python and R, improving code maintainability and efficiency for client projects.

# PROJECT WORK

- **Financial Navigator App (2025):** Engineered an AI agent-driven Streamlit application for real-time stock analysis (fundamentals, performance, charts) and personalized SIP planning. Leveraged LangChain and multiple LLM providers to deliver dynamic, easy-to-understand insights, significantly enhancing user financial decision-making.

- **Change Point Detection Project (2025):** Built an advanced change point detection system leveraging statistical, machine learning, and generative AI techniques with Large Language Models (LLMs) for insightful analysis. Integrated data preprocessing and real-time visualizations in a Streamlit app, providing users with AI-generated explanations of detected changes. This innovative tool empowers users to understand and act on time series shifts quickly, enhancing decision-making across domains.

- **Karen Virtual Assistant (2020-2022):** Developed a sophisticated voice-activated AI system encompassing over 30 functions, leveraging advanced Natural Language Processing (NLP) techniques, Python programming, Recurrent Neural Networks (RNNs), Speech Recognition, and Text Generation using libraries such as Keras and NLTK. Enhanced user accessibility by seamlessly integrating mobile input through a dedicated application, significantly improving usability across multiple platforms. Implemented dynamic user control with a live Graphical User Interface (GUI) using PyAutoGUI, ensuring intuitive interaction and real-time responsiveness. For the neural network architecture, Rectified Linear Unit (ReLU) was employed as the activation function, optimizing performance in feature extraction and nonlinear mappings. The model was trained using categorical cross-entropy loss function, tailored for multi-class classification tasks, ensuring robust training and prediction accuracy. Integrated functionalities include real-time updates on COVID-19 statistics such as cases, deaths, and recoveries based on user-specified state and region, enhancing public health awareness. Additional features incorporated are a news teller for current affairs updates, a PDF reader for document accessibility, an alarm system for timely reminders, and numerous other customizable functions, enriching user experience with multifaceted utility and convenience.

- **Lung Cancer Detection using Convolutional Neural Network (2020):** Utilizing a deep learning CNN architecture, our model is trained on a dataset of 5000 lung images across three classes: Normal, Lung Adenocarcinomas, and Lung Squamous Cell Carcinomas. The Sequential model includes three Convolutional Layers followed by MaxPooling, a Flatten layer, and two fully connected layers. BatchNormalization layers ensure stable training, and Dropout minimizes overfitting. The final Softmax output layer provides accurate predictions, achieving an f1-score above 0.90 for each class, demonstrating high accuracy in lung cancer detection with correct predictions 90% of the time.

# EDUCATION

| | |
|---|---|
| **B.E. in Computer Engineering with Honours in Data Science** | Aug 2018 - July 2022 |
| **RSCOE, Pune University** | CGPA : 9.6 |

Relevant Coursework: Object Oriented Programming, Databases, Data Structures and Algorithms, Machine Learning, Java.

# AWARDS AND CERTIFICATES

- **Performance Driver Award:** Recognized for rapid acquisition of predictive analytics skills, exceptional contribution to database configuration automation tool development, and commendable dedication. Acknowledged for ongoing work on ML-based forecasting algorithm for dust collector system.

- **WoW Award:** Awarded the prestigious Wow Award for outstanding contributions across various projects. Played a crucial role in improving system performance and ease of maintenance by enhancing code in the Fuel On Demand project. Created a sophisticated quarterly model report system and established reliable failure notifications using SNS logs. Developed advanced clustering models for silica assessment in bauxite and successfully implemented an anomaly detection model in the PLSR project. This award recognizes innovative approaches and commitment to excellence in data science and model development, demonstrating leadership and significant contributions to the field.

- **Generative AI for Everyone:** by DeepLearning.AI

- **Machine Learning Specialization :** by Stanford University.

- **Python Specialization :** by University of Michigan.