Maroti Raghunath Shelke

✉ marotishelke2000@gmail.com / ▌ +91 9075325154
⌨ https://github.com/Marotishelke /in https://www.linkedin.com/in/maroti-shelke

SUMMARY

Data Scientist with 2 years experience in predictive maintenance and spectral data analysis. Proficient in advanced machine learning and deep learning techniques for driving data-driven decisions.

SKILLS

Programming Languages: Python, R, C, SQL

Machine Learning & Deep Learning Algorithms: Linear Regression, Logistic Regression, Quadratic Regression, Random Forest, Naive Bayes, Decision Trees, KNN, Clustering, Bagging, Boosting, Market Basket Analysis, Anamoly Detection, Dimension Reduction, Feature Selection, Convolutional Neural Networks, Recurrent Neural Networks, Artificial Neural Networks, Time Series Forecasting, Prophet, Arima, Data Structure and Algorithms, Recommendation Systems, Large Language Models, Generative AI

Technologies & Tools: Django, Mircosoft SQL Server, Azure DevOps, Azure Machine Learning Studio, Azure Synapse Pipelines, Docker, TensorFlow, PyTorch, Scikit Learn,Pandas,Numpy, LangChain, GIT, PySpark.

WORK EXPERIENCE

Capgemini, Pune   Oct 2023 - Present

Data Scientist

- Developed an approach to analyzing silica content in bauxite involved initial preprocessing steps such as noise reduction (Savitzky-Golay), baseline correction (IRSQR), and normalization (MSC, SNV). Dimensionality reduction using UMAP facilitated clustering via Agglomerative Hierarchical and K-means methods, while PLSR enabled predictive modeling of silica content with validation metrics like recall and MSE. We also conducted anomaly detection by comparing spectra to reference using Euclidean, Cosine Similarity, and Minkowski distances, visualizing results through histograms and GIFs for bin-wise and median analyses. These methods collectively enhance spectral analysis for precise assessment of ore quality and composition.

- Developed machine learning models to predict truck refueling times and accurately estimate remaining fuel, increasing efficiency by 15 hours and optimizing truck tank capacity utilization. Implemented advanced feature engineering techniques and the kernel density algorithm to enhance model accuracy. Used Matplotlib and Plotly for insightful visualizations, enabling effective data interpretation and decision-making. Contributed expertise in machine learning, feature engineering, and data visualization using Python, scikit-learn, and TensorFlow to streamline logistics operations.

- Predicted mining asset health by forecasting machine behavior and collaborated with clients to pre-inform them about upcoming failures minimizing shutdowns and production losses.

- Created a time forecasting model using Prophet to forecast the health of dust collector bags, providing a 21-day forecast. This model enabled regular monitoring of the dust collector's health and allowed the extension of maintenance windows when bags still had remaining life based on model predictions. Achieved a model accuracy of 94%, reducing maintenance shutdowns from 6 to 2 per year.

Senior Analyst   Aug 2022 - Oct 2023

- Implemented Quadratic Regression model to detect impeller wear out failures by calculating Remaining useful life (RUL) and forecasting Risk of Failure three weeks in advance. Successfully extended impeller component lifespan by over one year compared to traditional maintenance practices. Achieved a model accuracy of 90 % for impeller condition forecast leading to enhanced operational efficiency and cost savings.

- Developed a customized web application using Django framework for conducting preliminary data analysis and data visualization efficiently. Users can easily upload datasets in various formats (CSV, Excel) through a user-friendly interface.This tool enable users to perform EDA, including summary statistics, visualizations (histograms, scatter plots, etc.), and correlation analysis. This app help users gain insights into their data distributions and relationships.

- Created a Django Application that assesses the quality of Python and R code and generates comprehensive reports detailing code quality and providing suggestions for code improvement to enhance efficiency and maintainability using Pylint and Rlint libraries.

PROJECT WORK

- **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.
- **YouTube Ad Views Prediction (2021):** Developed a predictive model using regression techniques to estimate YouTube ad views, employing advanced machine learning methodologies such as feature selection and extraction. Conducted thorough data preprocessing, including outlier detection and normalization, to ensure data quality and model reliability. Delivered a valuable tool for advertisers, enabling them to strategically optimize content placement and audience targeting based on predicted ad views, thereby maximizing campaign effectiveness and return on investment.
- **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.
- **Twitter Sentiment Analysis (2019):** Developed a dynamic Twitter Sentiment Analysis Web App using Django Framework, leveraging Tweepy for Twitter API integration and TextBlob for sentiment analysis. Engineered data parsing algorithms to extract and analyze tweet sentiments based on subjectivity and polarity, facilitating real-time sentiment insights for users.

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.