Simran Rajendra Chaudhari

[simran.r.chaudhari@gmail.com](mailto:simran.r.chaudhari@gmail.com) *|* +91 8956418115

 github.com/chaudharisimran *|* linkedin.com/in/chaudhari-simran/

# Summary

Data Scientist with 2 years experience in predictive maintenance, adept at driving data-driven decisions through advanced machine learning and deep learning techniques

# Skills

**Programming Languages:** Python,R,Java,SQL,JavaScript,HTML,CSS

**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 De- tection, Dimension Reduction, Feature Selection, Convolutional Neural Networks, Recurrent Neural Networks, Artificial Neural Networks, Time Series Forecasting, Data Structure and Algorithms, Recommendation Systems, Large Language Models, Generative AI

**Technologies & Tools:** Django, Flask, 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***

* Predicted mining asset health by forecasting machine behavior and collaborated with clients to pre-inform them about upcoming failures minimizing shutdowns and production losses.
* Developed a machine learning model using the XGBoost algorithm to predict the remaining useful life of the mantle liner in a gyratory crusher, allowing operators to proactively schedule maintenance activities, order replacement parts in advance, and optimize crusher performance. Achieved a maintenance cost reduction of $1.5M and decreased the number of maintenance shutdowns from 4 to 2 per year, with the model attaining an accuracy of 94.%.
* Created a machine learning 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 91%, reducing maintenance shutdowns from 4 to 2 per year.
* Developed a model to analyze silica content in bauxite ore. Initial Preprocessing steps involved noise reduction using Savitzky-Golay filter, baseline correction using IRSQR, normaliztion using MSC,SNV and dimensionality reduction using UMAP. Employed Agglomerative Hierarchical and K-means clustering algorithms to uncover inherent patterns and groupings within the data, facilitating insightful data exploration.Leveraged Partial Least Squares Regression (PLSR) for predictive modeling of silica content based on extracted features, enabling accurate estimation. These methods collectively enhanced spectral analysis for precise assessment of ore quality and composition.

***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

* **EmoBot AI-Powered Emotion Analysis and Music Recommendation Chatbot (2022):** Developed a chatbot designed for user interaction, emotion analysis, and personalized song playlist recommendations.The primary goal is to identify the user’s emotions conveyed through chat interactions. Based on the detected emotion, the chatbot suggests a curated list of songs that align with the user’s emotional state.Users have the ability to listen to and enjoy the recom- mended songs while continuing to chat with the chatbot. Implemented as a web application, ensuring accessibility and ease of use across different devices and platforms.Utilized a retrieval-based chatbot model that responds to users using predefined input patterns and corresponding responses.Implemented a specialized machine learning model, specifically a Long Short-Term Memory (LSTM) neural network, to classify the category of user messages effectively.Employed var- ious machine learning classifiers including Support Vector Machine (SVM), Linear SVM, Random Forest, and Decision Tree models to accurately detect and interpret the emotions expressed by users in their chats.
* **Movie Recommendation System (2021):** Implemented a Movie Recommendation System based on Content-based filtering strategy, employing cosine similarity calculations to suggest items aligned with user preferences. Engineered backend functionalities to analyze user preferences and item attributes such as lead actors, directors, and genres, ensuring accurate and personalized recommendations. Designed a visually appealing movie website frontend using Django, enhancing user experience and engagement with the recommendation system
* **Twitter Sentiment Analysis (2020):** 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.
* **Lung Cancer Detection using Convolutional Neural Network (2019):** Utilizing a deep learning architecture, specifically designed CNN models are employed. These networks are trained on annotated datasets consisting of both cancerous and non-cancerous lung images. The CNN’s layers, including convolutional, pooling, and fully connected layers, enable it to learn distinctive features and patterns indicative of lung cancer.This dataset includes 5000 images for three classes of lung conditions:Normal Class,Lung Adenocarcinomas,Lung Squamous Cell Carcinomas We have implemented a Sequential model which contains Three Convolutional Layers followed by MaxPooling Layers.The Flatten layer to flatten the output of the convolutional layer.Then we will have two fully connected layers followed by the output of the flattened layer. We have included some BatchNormalization layers to enable stable and fast training and a Dropout layer before the final layer to avoid any possibility of overfitting.The final layer is the output layer which outputs soft probabilities for the three classes.The performance of our simple CNN model is very good as the f1-score for each class is above 0.90 which means our model’s prediction is correct 90% of the time.

# Education

**B.E. in Computer Engineering** Aug 2018 - July 2022

**MMCOE, Pune University** CGPA : 8.9

**R**elevant Coursework: Object Oriented Programming, Databases, Data Structures and Algorithms, Machine Learning.

# Awards and Certificates

* + **Performance Driver Award:** Recipient of the Performance Driver Award for rapidly acquiring skills in Predictive Analytics and spearheading the development of standardized preliminary data analysis tools and ML models. Notable contributions include designing time series machine learning forecasting models.
  + **WoW Award:** Awarded the prestigious Wow Award for exceptional programming prowess and innovative development of a standard code checker tool for R and Python. Recognized for expertise in time series machine learning models and demonstrated potential as a proficient Data Scientist.
  + **Generative AI for Everyone:** by DeepLearning.AI
  + **Machine Learning Specialization :** by Stanford University.
  + **Deep Learning Specialization :** by Stanford University.
  + **Python Specialization :** by University of Michigan.