

# Rajesh S H

📍 India ✉ shrajesh.work@gmail.com ☎ 07338077439 📧 in/rajesh-s-h-592591211 🌐 shrajeshwork.wixsite.com/portfolio-of-1

---

## PROFESSIONAL SUMMARY

---

Experienced Data Scientist skilled in deploying deep learning models, optimizing resource allocation through forecasting, and automating data workflows across various domains.

---

## DATA SCIENCE EXPERIENCE

---

### Data Scientist | Diggibyte Technologies | Bengaluru | February 2024 - May 2024

- Experience in PySpark with Databricks, Large Language Model and Generative AI.
- Deep Learning Model for Object detection and Tracking with counts.
- Knowledge on Machine learning in Associate and professional levels with Databricks.
- Built Summarization model for Norway language article based on client requirement.
- Engaged in hackathon hosted by Databricks where we created a chatbot using RAG based LLM to reduce risks involved during manufacturing.
- In-depth experience in data visualization, EDA, profiling the data using Databricks. Experience working with Microsoft Azure cloud platform.

### Machine Learning Engineer | Freelance | New York, NY | July 2023 - January 2024

- Created extensively in the medical domain, focusing on protein abundance analysis in organisms.
- Developed a polynomial regression model achieving 92% efficiency, a significant increase from 67%.
- Implemented rigorous data cleaning, preprocessing techniques, and advanced data visualization to enhance insights.
- Utilized robust outlier detection techniques to ensure data integrity and model reliability.
- Implemented feature engineering strategies to enhance predictive performance and model interpretability.

### Data Scientist | Soothsayer Analytics | Bengaluru | February 2023 - April 2023

- Developed end-to-end pipeline for data processing, visualization, and feature engineering.
- Built predictive models with 95% accuracy, reducing errors by 30%.
- Optimized resource allocation through demand forecasting and customer behavior prediction.
- Leveraged tools like Matplotlib for dynamic data visualization, ensuring real-time updates and insights.
- Developed and deployed machine learning algorithms that increased prediction accuracy by 25%, analyzed large-scale datasets using Python, driving strategic business decisions based on data-driven insights.

### Data Scientist Trainee | TuringMinds.AI | Bengaluru | December 2021 - January 2023

- Developed deep learning segmentation model, increasing accuracy from 87% to 93% and reducing labeling errors by 40%.
  - Experience on projects in Medical Diagnosis and Flight Delay Prediction.
  - Practiced programming skills on Kaggle datasets.
  - Deployed advanced data augmentation techniques to enhance model generalization and performance.
  - Conducted exploratory data analysis to identify patterns and outliers in medical domain datasets.
- 

## EDUCATION

---

### Bachelor's of Computer Applications | Presidency College | Minor in Mathematics | Bengaluru | 2022 | 8.1

- Participated in IT fests, led teams, and won prizes
  - Organized and managed the 'GameX' IT fest event
- 

## PROFESSIONAL CERTIFICATIONS

---

### Scalable Machine Learning with Apache Spark™ | Databricks | 2024

### Computational Data Science | Internation School of Engineering(INSOFE) | 2023

### Computational Data Science | Great Learning | 2022

---

## SKILLS

---

Python, SQL, Machine Learning, Deep Learning, NLP, LLM, Generative AI, Model Development & Management, Data Analysis, Predictive Modeling, Statistical Analysis, Problem Solving, Decision Making, Microsoft Azure, Databricks, Excellent Communication, Multi-tasking

---

## COURSEWORK

---

### Computational Data Science | INSOFE | 2021

- Machine Learning
  - Deep Learning
  - Model Development & Management
  - Data Analysis
  - Python
- 

## PROJECTS

---

### AI-Based Resume Screening Tool | [github.com/Github-Rajesh/Github-Rajesh-AI-Based-Resume-Screening-Tool](https://github.com/Github-Rajesh/Github-Rajesh-AI-Based-Resume-Screening-Tool)

- Developed a web-based application utilizing machine learning to automate resume screening processes.
- The tool extracts text from PDF resumes using PyMuPDF, preprocesses the text, and compares it to job descriptions using a RandomForestClassifier model. Implemented the entire pipeline, including text preprocessing with NLTK, feature extraction with TF-IDF vectorization, and model training for predicting resume-job description matches.
- Designed and developed a Flask-based application for automated resume screening, optimizing candidate selection by integrating capabilities for uploading resumes and job descriptions.

### Conversational chatbot | Diggibyte Technologies | [github.com/Github-Rajesh/Chatbot-](https://github.com/Github-Rajesh/Chatbot-)

- Developed a web-based application that leverages AI to provide intelligent responses to user queries by analyzing PDF documents and recommending related YouTube videos.
- Utilized LangChain for document loading and splitting, OpenAI for embeddings and conversational retrieval, and Chroma for creating and persisting a vector database.
- Implemented an interactive chatbot interface using Flask, integrating OpenAI's GPT and YouTube Data API to enhance user interaction by fetching relevant video content based on the query context.

### Protein Abundance prediction | Freelance | [github.com/Github-Rajesh/Freelance](https://github.com/Github-Rajesh/Freelance)

- Developed a robust machine learning model to predict protein abundance using polynomial regression.
- The process involved loading and cleaning data from Excel, handling missing values, applying robust scaling to features, and removing unnecessary columns.
- Outliers detection using Z-score analysis. Polynomial feature transformation being applied, followed by training a linear regression model on the processed data.
- The model evaluation done using  $R^2$ , MAE, and RMSE, achieving high predictive accuracy with a strong  $R^2$  score. Additionally, a regression plot created to visualize the actual vs. predicted values, demonstrating the model's effectiveness.

### WEBCrawler | [github.com/Github-Rajesh/Listed-Assessment](https://github.com/Github-Rajesh/Listed-Assessment)

- The provided Python script efficiently retrieves YouTube links related to a specified query using Google search restricted to youtube.com.
  - The script optimally saves these links into both JSON and CSV formats using a unified 'save\_to\_file(data, output\_file, file\_format)' function, ensuring flexibility and maintainability.
  - This approach consolidates file-saving operations while enhancing clarity and error management, making it a practical solution for retrieving and storing YouTube links programmatically based on specific search criteria.
-