

# Suraj Rajolad

Data Scientist

+1(508) 509-0180 | [srajolad@umassd.edu](mailto:srajolad@umassd.edu) | [LinkedIn](#) | [GitHub](#) | [Tableau](#) | [Portfolio](#)

## TECHNICAL SKILLS

**Programming Languages & Scripting:** Python, R, Java, C, C++, VB, Shell, UNIX, SQL

**Data Science & Machine Learning Tools:** PyTorch, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, Flask, SciPy, ggplot, Mathplotlib, Time Series, TensorFlow, NLTK, Deep Learning, Stat models, OpenCV, NLP, KNN, SVM, KMeans, Random Forest, Supervised & Unsupervised Learning, Predictive Analytics, Reinforcement Learning, Transfer Learning, LLMs, Linear Regression, Decision Tree, ARIMA, SARIMA, Keras, RNN, MLflow, Airflow, Natural Language Processing, Computer Vision, Image Processing, Selinium, GenAI.

**Data Analysis & Visualization:** Exploratory Data Analysis (EDA), Feature Engineering, Data Visualization, Data Cleaning and Preprocessing, PowerBI. Tableau, Advanced Excel.

**Mathematics:** Advanced Mathematical Statistics (Statistical Modeling, Hypothesis Testing, Reg Analysis), Linear Algebra.

**Big Data Technologies:** Hadoop, HDFS, MapReduce, Kafka, Hive, Avro, Oozie, Scala, Machine Learning, Statistical Modeling, Predictive Models, Exploratory Data Analysis

**Cloud & DevOps:** AWS (EMR, Step Functions, S3, Lambda, CloudWatch, IAM, Sagemaker), Docker, Kubernetes, Tableau

**Databases:** SQL, MongoDB, CassandraDB, Snowflake, Geospatial Analytics

**Tools and Platforms:** JIRA, Git, Postman, PostgreSQL Git, GitHub, Docker, CI/CD pipelines, NoSQL

## EDUCATION

<b>MSc Data Science</b> (GPA: 3.6/4.0)	Sep 2023 - May 2025
University of Massachusetts Dartmouth	Dartmouth, MA
<b>Bachelor of Engineering</b> (GPA:7.6/10.0)	Aug 2017 - Jun 2021
Acharya Institute of Technology	Bengaluru, India

## WORK EXPERIENCE

<b>Workday Analyst</b>	Jul 2021 - Aug 2023
Wipro Limited	Bengaluru, India
<ul style="list-style-type: none"><li>Data Analysis &amp; Trend Identification: Analyzed trends in Service Cloud requests while working with the Meta team, uncovering the root causes of recurring system issues. By implementing machine learning-focused improvements, I helped reduce incidents by 30% within just three months, enhancing overall system reliability.</li><li>Workday Data Integrity: Led quarterly audits of the EIB data in Workday for Meta, uncovering discrepancies that were causing payroll inaccuracies. My actions directly contributed to resolving key issues, improving payroll accuracy, and ensuring smoother operations for the team.</li><li>Operational Efficiency: Created and rolled out new python-based protocols to eliminate duplicated information in the system, streamlining processes and reducing errors. This initiative saved the team about five hours every week, freeing up valuable time that could be spent on more impactful tasks.</li></ul>	

## PROJECT EXPERIENCE

### Multimodal AI with OpenAI CLIP [\[Link\]](#)

- Leveraged OpenAI's CLIP (Contrastive Language-Image Pre-training) model for zero shot image classification, aligning visual and textual embeddings using cosine similarity on a custom subset of the COCO dataset.
- Built interactive visualizations including similarity heatmaps and object detection bar charts to showcase model predictions; evaluated robustness to distribution shift and diverse real-world contexts using python and data visualization libraries.
- Utilized PyTorch, NumPy, pandas, and Matplotlib in a Google Colab notebook environment to implement and analyze CLIP's performance on multimodal tasks, demonstrating advanced deep learning and computer vision integration.

### Optimizing-the-Traveling-Salesman-Problem-Using-Genetic-Algorithms [\[Link\]](#)

- Implemented multiple Genetic Algorithm (GA) solutions for the Traveling Salesman Problem (TSP) using graph theory and coordinate-based distance metrics, optimizing route efficiency through elitism, mutation, and order-1 crossover techniques.
- Developed a custom Graph class in python to manage adjacency matrices, weighted edges, and dynamic path cost evaluation, enabling flexible experimentation with various TSP scenarios, leveraging pandas and numpy for data manipulation.
- Used numpy, matplotlib, and python's random module to visualize and validate optimized routes, incorporating tournament selection, predictive models for generation adjustments, and fitness-based optimization across multiple generations.

### Machine Learning Algorithms [\[Link\]](#)

- Implemented and visualized machine learning models including Linear Regression, Multiple Linear Regression, Logistic Regression, Decision Trees, and Support Vector Machines (SVM) using scikit-learn, Pandas, and Matplotlib along with numpy for computational operations.
- Applied techniques such as exploratory data analysis, preprocessing with SimpleImputer and StandardScaler, and evaluated models using metrics like R<sup>2</sup>, MAPE, MSE, Jaccard Index, confusion matrix, and classification reports.
- Developed interactive Jupyter Notebooks and Python scripts to analyze regression/classification tasks with visual insights including actual vs. predicted plots, scatter plots, and decision tree diagrams, and integrated outputs into PowerBI for holistic data visualization.

### Wage Prediction and Analysis using R [\[Link\]](#)

- Built a full data analysis and machine learning pipeline in R for wage prediction using Decision Trees and Lasso Regression, applying libraries such as caret, glmnet, and rpart for statistical modeling and performance evaluation.
- Conducted extensive data preprocessing and feature engineering using dplyr, tidyr, and pandas, including scaling, dummy encoding, and handling missing values to ensure data readiness for regression analysis and classification tasks.
- Visualized data distributions and model performance using ggplot2, GGally, and PowerBI, showcasing insights through histograms, boxplots, ROC curves, confusion matrices, and feature importance plots; evaluated predictive models using RMSE, R<sup>2</sup>, and ROC-AUC metrics.

### Accelerating Predictive Analytics on Large Datasets through Embarrassingly Parallel Computing. [\[Link\]](#)

- Implemented a predictive modeling pipeline using GradientBoostingClassifier and scikit-learn to classify book genres on a large dataset (100,000+ rows), applying data normalization, feature selection, and cross-validation to ensure model efficiency.
- Accelerated computation using embarrassingly parallel processing with Python’s multiprocessing library and the Pool() module, achieving up to 15x speedup and 60%+ efficiency on multi-core systems (6, 12, and 24 cores).
- Conducted detailed performance benchmarking comparing serial and parallel executions, visualizing results through PowerBI and matplotlib to analyze trade-offs between data size, execution time, and resource utilization.

### Temporal Insights in Hospitality [\[Link\]](#)

- Applied ARIMA (Auto-Regressive Integrated Moving Average) modeling to forecast hotel occupancy rates in Boston using economic indicators from 2013–2019; performed seasonal analysis and identified peak months for profit optimization in predictive models.
- Conducted comprehensive data preprocessing and visualization using python, seaborn, and matplotlib; created correlation heatmaps, time series plots, and used ACF/PACF to guide parameter selection with regression analysis.
- Achieved meaningful predictions with low RMSE, showing July as the peak month with ~90% occupancy; recommended dynamic pricing strategies based on data trends and SQL exploration from Logan passenger traffic reports.

### IBM-Data-Science-Capstone-SpaceX [\[Link\]](#)

- Collected and wrangled Falcon 9 launch data from SpaceX REST API and Wikipedia to engineer features and predict first-stage landing success using python-based machine learning classification models.
- Conducted exploratory data analysis with SQL, Matplotlib, Folium, and Plotly Dash, revealing insights like optimal launch sites, orbits with 100% success, and the effect of payload mass on landing outcomes.
- Built and tuned Logistic Regression, SVM, Decision Tree, and KNN predictive models using GridSearchCV in scikit-learn, identifying Decision Tree as the best performer with the highest cross-validated score.

### Tableau Dashboards and Visual Analytics [\[Link\]](#)

- Developed interactive Tableau dashboards across domains like airline performance, retail sales, European markets, and online course analytics, enabling data-driven decision-making through dynamic data visualization storytelling enhanced with matplotlib and numpy.
- Integrated SQL filters, KPIs, geographic mapping, and trend analysis to deliver actionable insights on flight delays, regional sales performance, and course popularity using diverse datasets.
- Designed quiz-based and domain-specific dashboards focused on comparative analysis, supported by exploratory data analysis, showcasing adaptability across business use cases.

## COURSES

**Tableau 2024 A-Z: Hands-On Tableau Training for Data Science, Udemy.**

**IBM Data Science Professional Certificate.**

**Artificial intelligence A-Z 2025, Udemy**