

# MUTHARASAN M

## DATA SCIENTIST

### CONTACT

+91 8072686269  
mutharasanmanoharan@gmail.com  
<https://github.com/Muthu2312>  
Saidapet, chennai

### SKILLS

- C, C++, Python, R ( Basic )
- DBMS ( Mysql, Mango DB, etc )
- Statistics
- Tableau, Power BI
- Django, Flask, Streamlit
- Html, CSS, JavaScript ( Basic )
- Machine Learning
- Deep Learning ( Neural Networks )
- Computer Vision ( Opencv, Yolo )
- Natural Language Processing
- Transformers ( LLM, Bert, etc )
- GenAi
- Git / Github
- Docker / Kubernetes
- Cloud Computing ( AWS )

### EDUCATION

#### PGPDSE-DATA SCIENCE

THE GREAT LAKES INSTITUTE OF  
MANAGEMENT (2021-2022)

#### B.E MECHANICAL ENGINEERING

VSU COLLEGE OF ENGINEERING  
TECHNICAL CAMPUS (2013 - 2017)

### LANGUAGES

Tamil

Engilsh

### PROFILE

Data Scientist with three years of experience in data collection, analysis, and modeling. Proven ability to develop predictive models, perform exploratory data analysis, and deploy scalable solutions using cloud platforms. Skilled in collaborating with cross-functional teams to deliver data-driven insights and solutions. Proficient in deploying models using MLOps and DevOps best practices.

### WORK EXPERIENCE

#### Data scientist

Spiro prime tech services 2021-2024

##### Data Collection and Processing

- Data Gathering: Collect data from various sources, including databases, APIs, and external datasets.
- Data Cleaning: Handle missing data, remove duplicates, correct inconsistencies, and preprocess raw data for analysis.

##### Data Analysis and Interpretation

- Exploratory Data Analysis (EDA): Perform initial investigations on data to discover patterns, spot anomalies, and test hypotheses.
- Statistical Analysis: Apply statistical methods to understand data distributions and relationships among variables.

##### Modeling and Evaluation

- Model Building: Develop predictive models using machine learning algorithms.
- Model Evaluation: Validate models using metrics like accuracy, precision, recall, and AUC-ROC. Fine-tune models for optimal performance.

##### Communication and Reporting

- Data Visualization: Create visual representations of data findings using tools like Matplotlib, Seaborn, or Tableau.
- Reporting: Prepare and present reports summarizing insights, findings, and recommendations to stakeholders.

##### Collaboration

- Cross-functional Collaboration: Work with other teams (e.g., engineering, product, marketing) to understand business needs and integrate data solutions.
- Big Data Tools: Utilize big data tools like Hadoop, Spark, or cloud services (e.g., AWS, Azure) for large-scale data processing.

##### Deployment

- Deployment: Deploy models and applications using frameworks such as Django, Flask, or REST API to make them accessible to users.

##### DevOps and MLOps

- Continuous Integration (CI): Implement CI pipelines to automatically test and integrate code changes using tools like Jenkins, GitLab CI/CD, or GitHub Actions.
- Continuous Deployment (CD): Set up CD pipelines to automatically deploy code changes to staging or production environments. Ensure seamless deployments and rollback strategies.
- Monitoring and Logging: Implement monitoring and logging to track performance and health of deployed models and applications using tools like Prometheus, Grafana, ELK Stack, or CloudWatch.
- Model Versioning and Registry: Use model versioning tools like MLflow, DVC, or AWS SageMaker Model Registry to track different versions of models and manage their lifecycle.
- Automated Model Retraining: Set up pipelines for automated model retraining based on new data or model performance degradation using orchestration tools like Kubeflow.
- Collaboration and Reproducibility: Use collaboration tools like JupyterHub or Google Colab for collaborative development. Ensure reproducibility of experiments and results using tools like Jupyter Notebooks, Docker, or Conda.
- Cloud Integration: Utilize cloud platforms for scalable and cost-effective deployment, storage, and processing solutions.