

MD ARSHAD HUSSAIN

Bangalore

+91 8877020394

✉ arshadattari007@gmail.com [in](#) [LinkedIn](#) [G](#) [Github](#)

Education

Ramgarh Engineering College

2021 – 2024

B.Tech in Computer Science and Engineering(Lateral Entry)

Ramgarh, Jharkhand

AL-Kabir Polytechnic, Jamshedpur

2018 – 2021

Diploma in Computer Science Engineering

Jamshedpur, Jharkhand

Government High School

2016 – 2017

Matriculation(State Board)

Koderma, Jharkhand

Relevant Coursework

Data Science — Python Programming — Database Management — Machine Learning
Data Visualization — SQL & NoSQL Databases — Statistics for Data Science

Experience/INTERNSHIP

RUBIXE

Feb 2025 – Aug 2025

Data Science Intern

Bangalore, Karnataka

- Working on real-world data science projects with a focus on Python, machine learning, and analytics.
- Handling data preprocessing, model building, and performance evaluation under industry mentors.
- Utilizing tools such as Pandas, Scikit-learn, and Power BI for data-driven decision making.

Projects

Insurance Cost Prediction | *Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn*

[GitHub](#)

- Built a regression model to predict individual medical insurance costs using features like age, BMI, smoking status, and region.
- Applied preprocessing, feature encoding, and outlier detection to improve model reliability.
- Visualized feature relationships to identify cost drivers and trends.
- Compared Linear Regression, Ridge, Lasso, and Decision Tree models — **improved RMSE by 15% over baseline model.**

Flight Fare Prediction | *Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn*

[GitHub](#)

- Developed a predictive model for flight ticket prices using airline, source, destination, and duration features.
- Conducted EDA and data cleaning to extract meaningful patterns.
- Trained and tuned Random Forest and XGBoost — **achieved 92% accuracy on test data, reducing error rate by 18% compared to linear models.**

INX Future Inc (IABAC Project) | *Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn*

[GitHub](#)

- Designed classification models to predict employee performance levels based on department, education, and training hours.
- Performed feature engineering and analysis to highlight key drivers of performance.
- Implemented Decision Tree and Random Forest — **increased F1-score by 10% compared to baseline classifier.**

Technical Skills

Languages: Python, SQL

Tools & Visualization: Power BI, Google Colab, Jupyter Notebook, VS Code, Git/GitHub

Libraries & Frameworks: Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow, Matplotlib, Seaborn, Plotly

Certifications

- Certified Data Scientist- Datamites
- Python Programming - Naresh IT
- Certified Data Scientist- IABAC
- Certified Data Scientist- NASSCOM