



shoaibsayed687@gmail.com
7385746998
pune

Sayed Shoib

DATA SCIENTIST

SKILLS

Programming & Databases

- Python, SQL, Postgresql

Python Libraries

- NumPy, Pandas
- Scikit-Learn
- TensorFlow, PyTorch
- Matplotlib, Seaborn
- NLTK, SpaCy, Transformers (HuggingFace)

ML & DL Tools

- Anaconda (spyder)
- Jupyter Notebook
- Google Colab
- OpenCV
- MLflow

Data Visualization

- Power BI
- Tableau
- Excel (Pivot Tables, Power Query)

Version Control

- Git, GitHub

Cloud & Deployment (Optional)

- AWS (S3, Lambda, EC2)
- Flask / FastAPI

CORE COMPETENCIES

- Data Science
- Machine Learning
- Data Regression Analyses
- NLP/Advance
- NLP Data Mining
- Data Preprocessing
- Data Cleansing,
- Modeling
- Reinforcement Learning
- Predictive Modelling
- Deep Learning using CNN, RNN, LSTM, BERT and Transformer
- Stakeholder Management

CERTIFICATIONS

Oracle Cloud Infrastructure 2025 Certified Data Science Professional

Issuer :- ORACLE

Quantitative Research Job Simulation JP MORGAN CHASE & CO

Issuer :- FORAGE

PROFILE

Data Analytics & Machine Learning fresher with practical experience in building dashboards, applying statistical techniques, and developing ML models to support business decision-making. Completed projects in customer analytics, sales forecasting, fraud detection, and customer segmentation. Proficient in Python (Pandas, NumPy, Scikit-Learn), SQL, Excel, and data visualization. Strong understanding of business KPIs in marketing, finance, operations, and e-commerce. Passionate about creating data-driven impact.

EDUCATION

B-TECH CGPA-7.1

Maharashtra Institute Of Technology Aurangabad

2022-2026

HSC percentage-80

PODAR International Latur

2019-2021

SSC percentage-90.60

B.W.NEW English Medium School Latur

2018-2019

PROJECTS

Uber Fare Prediction Model

The Uber Fare Prediction Model is a machine learning system that estimates the cost of a ride based on historical trip data. It uses features like pickup and drop-off locations, distance, time of ride, and passenger count to predict fares accurately. Data preprocessing involves cleaning, outlier removal, and feature engineering such as calculating trip distance using the Haversine formula. The model is trained using regression and ensemble methods like Linear Regression, Random Forest, or XGBoost. It is evaluated using metrics such as RMSE and MAE to ensure accuracy. This model helps provide real-time fare estimates, supports surge pricing strategies, and enhances the overall user experience in ride-hailing platforms

Crop Disease Detection Model

Developed an AI-powered system combining deep learning, computer vision, and web-based monitoring for real-time disease detection and automated spraying. Trained and optimized a CNN model for edge deployment, integrated OpenCV for image processing, and built a Flask backend with interactive mapping. Utilized NumPy, Pandas, and visualization tools for performance analysis and ensured low-latency, responsive decision-making. Skills used Python, CNN, Tensorflow, Keras, Flask, Flutter, Computer Vision, Image Processing.