

# Unemployment Rate Prediction: Data Science Summary

Objective: Analyze and predict unemployment trends in India using historical data.

Dataset: Unemployee.csv

Key Findings from EDA:

- Unemployment rose post-COVID, especially in rural regions.
- Distribution is right-skewed, concentrated between 10-15%.
- Rural areas show greater volatility and higher rates.
- Regional disparities exist in unemployment levels.

Outlier Handling: IQR method used to clean data.

Feature Engineering:

- Added COVID period indicator.
- Categorical encoding and feature selection applied.

Model Comparison:

- Lasso:  $R^2 \sim 0.076$
- Ridge:  $R^2 \sim 0.078$
- Random Forest (Tuned):  $R^2 = 0.6523$

Conclusion:

- Random Forest shows strong predictive capability.
- Suitable for deployment and real-time use.
- SARIMA supports long-term trend analysis.

Report Generated using ReportLab - Python