BACHELOR THESIS PROJECT

TITLE

GDDP PREDICTION USING NIGHT-TIME LIGHT DATA FOR INDIA

PRESENTED BY : SANKET JAGTAP (21AG10032)

SUPERVISOR: PROF. ANUBHAB PATTANAYAK

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

ABSTRACT

Goal

To predict district-level GDP using night-time light (NTL) data, offering near-real-time economic insights

Approach

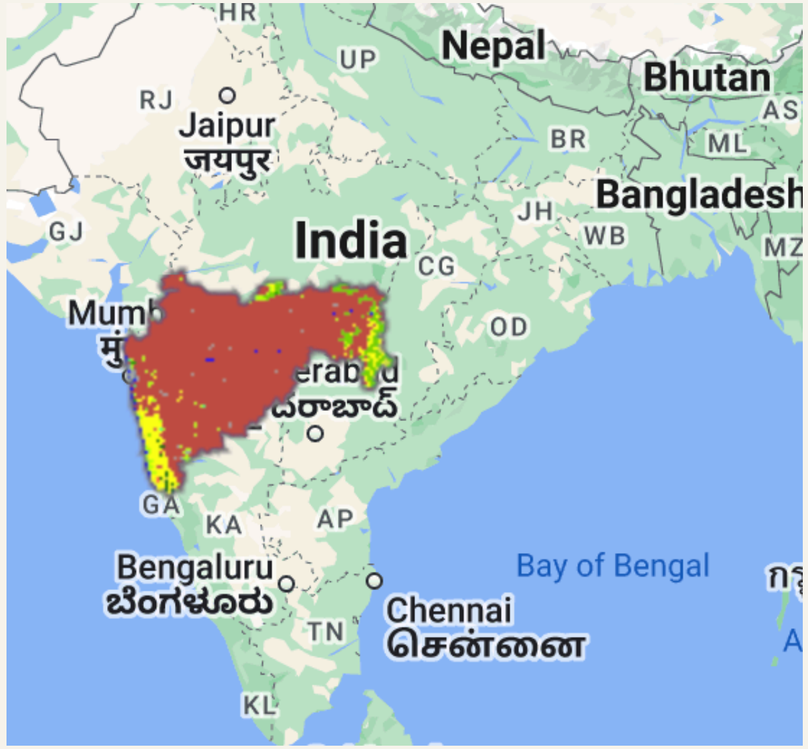
Utilized NTL data as a proxy for economic activity, incorporating machine learning (ML) and deep learning (DL) models to predict GDP

Significance

Provides a scalable method for assessing economic conditions, helping policymakers in resource allocation and decision-making

INTRODUCTION

- Context: GDP is crucial for economic analysis, but traditional data collection is slow.
- Solution: Use of satellite-captured NTL data as an alternative to infer economic activity, as regions with higher GDP tend to have greater night-time light intensity.
- Advancements: This approach leverages data analytics and satellite imagery, enabling a more responsive understanding of economic trends.



PROBLEM

Challenge

Traditional GDP data collection is limited by time lags and data accessibility, especially in under-reported areas

Objective

1) Develop a predictive framework to estimate GDP using NTL data
2) Predictiong the GDDP data from year 1992 to year 2022

OBJECTIVES

Primary Goal

Predict district-level GDP in India using night-time light data from year 1992 to year 2022 on the basis of available NTL data from 1999 to 2013

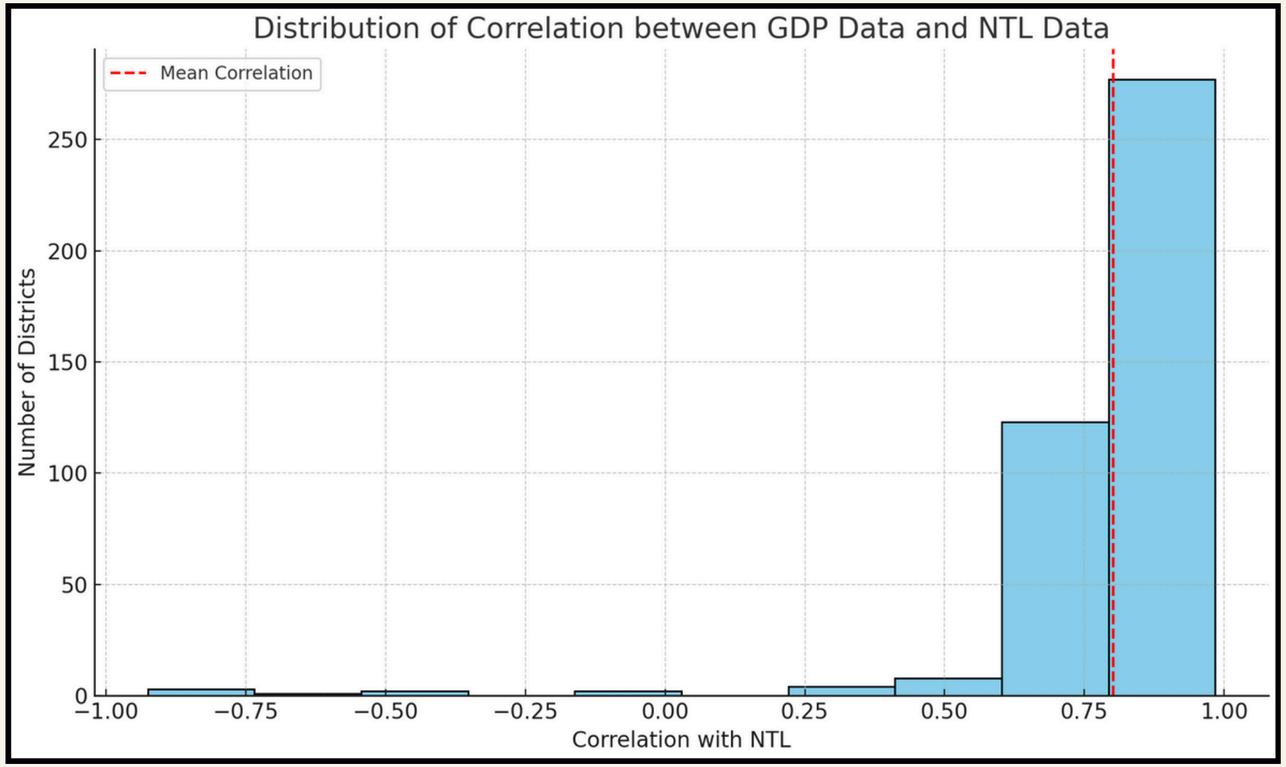
Modelling Aim

Develop a model capable of predicting GDP accurately in years before 1992 (backcasting) and after 2022 (forecasting)

HYPOTHESIS

Main Hypothesis: NTL data can effectively forecast GDP by capturing economic activity across districts, serving as a proxy in areas with limited traditional data

Correlation between GDP data and NTL data: 0.76



METHODOLOGY

Data Collection

Scraping and integrating NTL data from DMSP and VIIRS with district-level GDP data

Model Selection

Multiple ML algorithms, such as Gradient Boosting and Linear Regression, were tested

Feature Engineering

Included annual NTL features and scaled data for better interpretability

Extrapolation

Used NTL data from 1992 to 2022 to predict GDP values beyond available years

IMPLEMENTATION

Phase 1

Data collection and cleaning (From Google Earth Engine, ICRISAT, Planning Commission)

Phase 2

Feature engineering and model training with NTL and GDDP data

Phase 3

Model testing, selection, and evaluation of best-performing models

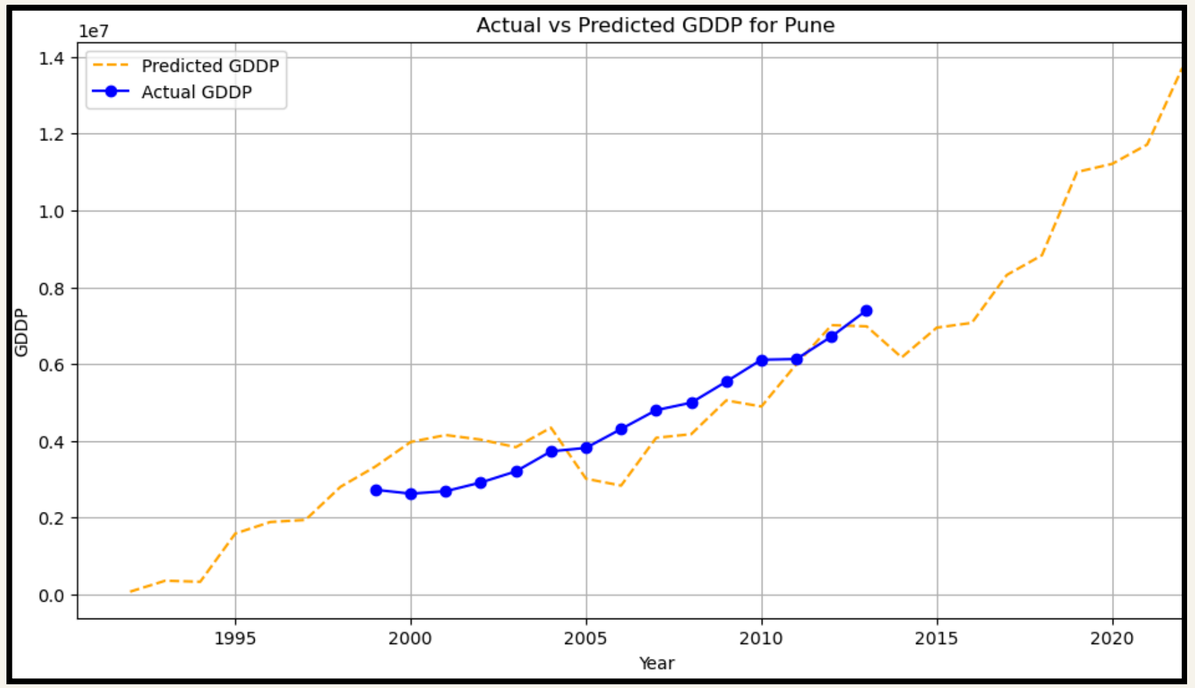
Phase 4

Final implementation and evaluation on unseen data

RESULT

• Outcome: The final model achieved high accuracy in predicting GDP, with Linear Regression performing best due to its balance of simplicity and robustness

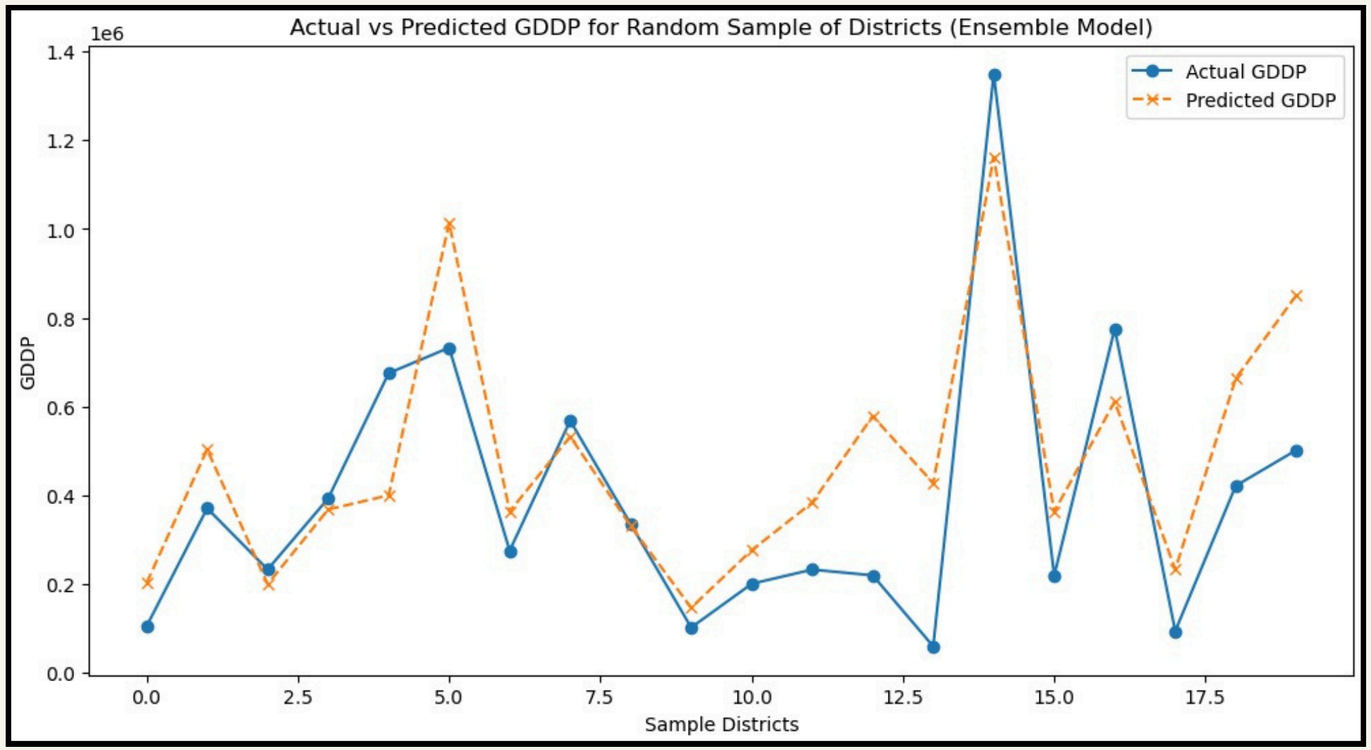
• Key Metrics: R^2 of approximately 0.89 for the Linear Regression model, indicating a strong fit between NTL and GDDP data



CONCLUSION

- Achievement: Successfully demonstrated that NTL data is a reliable proxy for GDP estimation, especially in under-reported districts.

- Policy Implications: This model enables more responsive economic planning and resource allocation for regional growth



RECOMMENDATION

- **Future Models**

  Consider testing advanced models like XGBoost with more data to further improve predictions

- **Expanded Features**

  Integrate additional regional socio-economic data to capture more complex economic patterns

THANK YOU

PRESENTED BY : SANKET JAGTAP