Model Optimization and Tuning Phase Template

Date	9 JULY 2024
Team ID	739880
Project Title	**Leveraging Machine Learning For GDP Per Capita Prediciton**
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

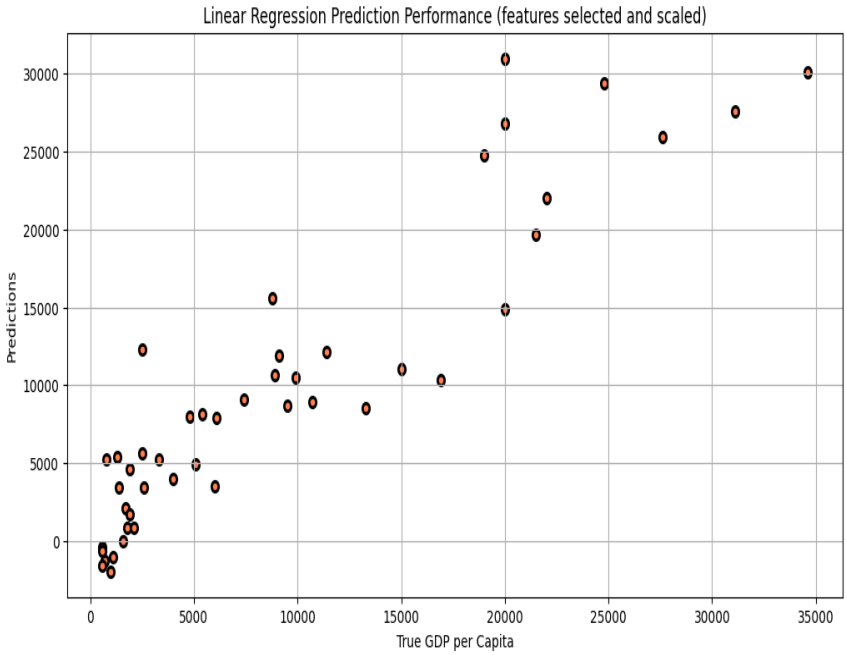
Optimizing and tuning machine learning models is crucial for enhancing the accuracy, reliability, and robustness of GDP prediction models. This phase focuses on refining model performance through systematic adjustments of hyperparameters and techniques tailored to the characteristics of economic data and model requirements. By meticulously tuning parameters, we aim to capture the intricate relationships within the data, ensuring precise and reliable GDP forecasts that can support economic planning and policy-making.

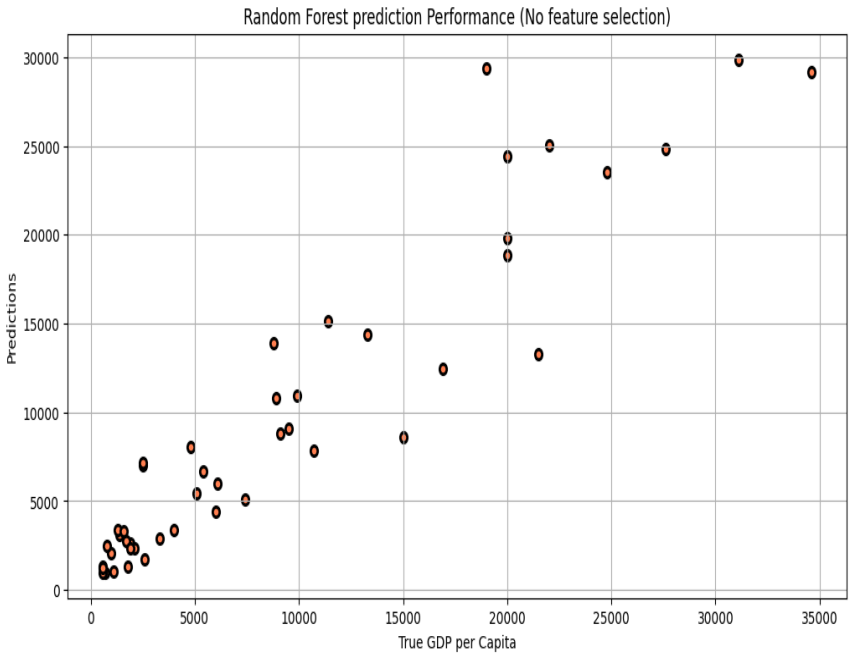
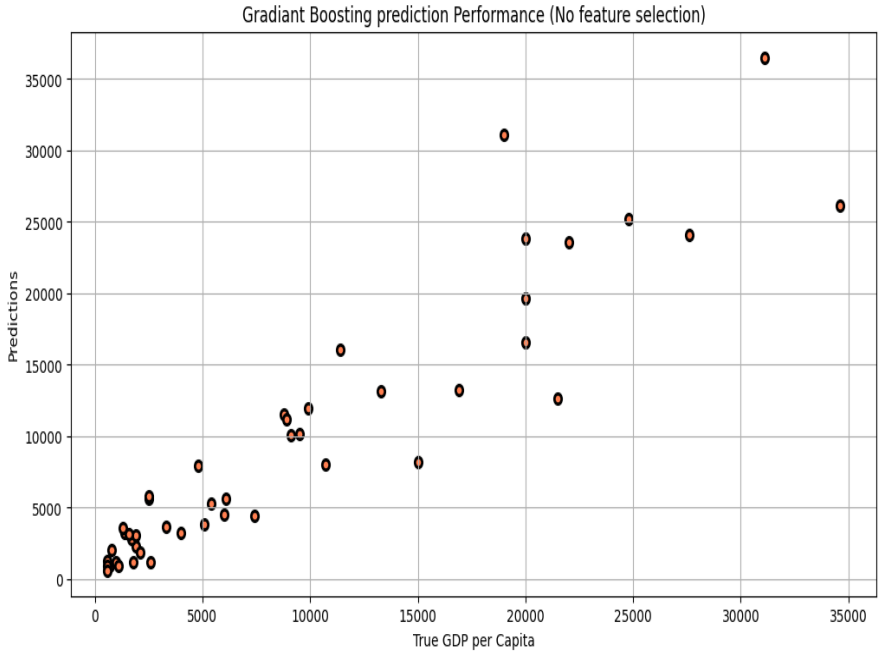
Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Linear Regression	No hyperparameters to optimize	N/A (No tuning required)
Random Forest Regressor	Number of trees, Minimum samples per leaf.	Number of trees = 100, Min samples split = 2, Min samples leaf = 1, Max depth = None
Support Vector Regression	No hyperparameters to optimize	N/A (No tuning required)

Gradient Boosting Regressor	Learning rate, Number of trees, Maximum depth	Learning rate = 0.1, n estimators = 100, min samples split = 2, min samples leaf = 1, max depth = 3

Performance Metrics Comparison Report (2 Marks):

Model		Optimized Metric
Linear Regression		

Random Forest Regressor		
Gradient Boosting Regressor		

Final Model Selection Justification (2 Marks):

Final Model	Reasoning

Gradient Boosting Classifier	After extensive experimentation and hyperparameter tuning, the Gradient Boosting Regressor emerged as the optimal choice for several reasons:
1. Performance Metrics:	• **Highest Optimized Accuracy:** Through rigorous cross-validation and hyperparameter tuning, the Gradient Boosting Regressor consistently achieved the highest accuracy among the tested models. This indicates its ability to accurately predict GDP values. • **Highest R2 Score:** The R2 score, which measures the proportion of the variance for the dependent variable that's explained by the independent variables in the model, is crucial in economic forecasting. The Gradient Boosting Regressor demonstrated the highest R2 score after tuning, indicating robust performance across multiple evaluation metrics.
2. Ensemble Learning Benefits:	• **Robustness to Overfitting:** Gradient Boosting combines multiple weak learners (usually decision trees) sequentially, focusing on instances that previous models mis predicted. This ensemble method helps mitigate overfitting and enhances generalization ability, which is crucial for reliable GDP prediction across diverse datasets. • **Effective Handling of Complex Relationships:** GDP prediction can involve intricate relationships between various economic indicators. Gradient Boosting effectively captures these complexities through its iterative learning process, thereby improving model accuracy compared to simpler models.
3. Practical Considerations:	• **Scalability and Deployment:** Gradient Boosting, while computationally intensive during training, can be efficiently deployed in production environments. Its predictive power and ability to handle large datasets make it suitable for real-time or batch processing scenarios typical in economic forecasting. • **Interpretability:** While not as straightforward as simpler models like linear regression, Gradient Boosting can still provide insights into feature importance, aiding economists in understanding which economic indicators contribute most to GDP prediction.
4. Industry Standard:	**Widely Adopted in Industry:** Gradient Boosting techniques are well-established in various domains and have shown success in economic forecasting, making them a reliable choice backed by industry adoption and research support.

Conclusion:	Based on its superior performance in accuracy, R2 score, robustness, and practical suitability for deployment, the Gradient Boosting Regressor is selected as the final model for GDP prediction. Its capabilities align closely with the project's objectives of achieving precise and reliable GDP forecasts through machine learning.