**Report**

**Title: *Big Data Solutions to the Affordable Housing Crisis in Urban Areas***

**Topic:** Smart City - Housing Affordability

**Subject code:** DS-670-HYB2-23WNTR

**Subject:** Capstone: Big Data & Bus Analy

**Professor:** Reda Mastouri

19/12/2023

**Team Members**

1. Saikiran Reddy Durgareddygari
2. Guna Venkata Sai Bharath Bodepudi;
3. Sahil Vericherla.

**Model Evaluation Report**

**Introduction:** This report provides an evaluation of machine learning models designed to address the complex challenge of housing affordability in urban areas. Leveraging smart city technologies, data-driven insights, and machine learning applications, the project aims to revolutionize the landscape of housing affordability.

**Data Preprocessing:**

* Imputation of missing values in 'Home\_Price,' 'Rental\_Price,' 'GDP\_Per\_Capita,' and 'Commute\_Time.'
* Feature Engineering: Created 'Housing\_Affordability\_Index' and 'Affordable' based on income, home price, and a threshold.

**Resampling Techniques:**

* Applied Synthetic Minority Over-sampling Technique (SMOTE) and Random Under-sampling to address imbalanced data.

**Models Evaluated:**

1. **Logistic Regression:**
   * AUC/ROC, Precision, Accuracy, MAE, MSE, RMSE, MAPE.
2. **Random Forest:**
   * AUC/ROC, Precision, Accuracy, MAE, MSE, RMSE, MAPE.
3. **Support Vector Machine (SVM):**
   * AUC/ROC, Precision, Accuracy, MAE, MSE, RMSE, MAPE.
4. **Gradient Boosting:**
   * AUC/ROC, Precision, Accuracy, MAE, MSE, RMSE, MAPE.

**Observations:**

* Logistic Regression:
  + Achieved results in AUC/ROC, Precision, and Accuracy, making it a promising candidate.
* Random Forest:
  + Demonstrated competitive performance across multiple metrics, indicating its effectiveness.
* Support Vector Machine (SVM):
  + Performed well with kernel='linear,' showcasing its potential for linearly separable data.
* Gradient Boosting:
  + Showed promising results in terms of AUC/ROC, Precision, and Accuracy.

**Next Steps:**

* Fine-tuning hyperparameters for optimal performance.
* Experimenting with additional algorithms and ensemble methods.
* Incorporating insights from peer-reviewed articles for enhanced model sophistication.