**Optimizing Smart City Public Transportation Through Data and Technology**

**Assignment 3**

**DS-670-HYB2-23WNTR**

**Capstone: Big Data & Bus Analysis**

**Reda Mastouri**

**19/12/2023**

**Team Members:**

Kethana devi kadiyala

Venkata Sai ram Rohith Mechineni

Deva Sai Vikas Vakkalagadda

**Smart City Public Transportation Project: Regression Model Evaluation**

**Introduction:** The evaluation of regression models is a crucial step in optimizing the Smart City Public Transportation Project. Focused on estimating travel time, this report presents the approach, models, and key metrics used to assess model performance.

**Data Preprocessing:**

* Handling missing values in 'Special\_Schedules.'
* Engineering 'Total\_Stops' to represent the total number of stops on a route.
* Splitting the dataset into features (X) and the target variable 'Estimated\_Travel\_Time' (y).

**Model Evaluation Framework:**

* Key regression metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), R-squared, and Explained Variance.

**Random Oversampling:**

* Addressing potential class imbalance by applying random oversampling to the training set.

**Regression Models:** Four models evaluated:

* Lasso Regression
* Decision Tree Regression
* Ridge Regression
* XGBoost Regressor

**Model Evaluation Results:**

* **Lasso Regression:**
  + MAE, MSE, RMSE, MAPE, R-squared, Explained Variance.
* **Decision Tree Regression:**
  + MAE, MSE, RMSE, MAPE, R-squared, Explained Variance.
* **Ridge Regression:**
  + MAE, MSE, RMSE, MAPE, R-squared, Explained Variance.
* **XGBoost Regressor:**
  + MAE, MSE, RMSE, MAPE, R-squared, Explained Variance.

**Observations:**

* Insightful evaluation of model performance in estimating travel time for public transportation.

**Next Steps:**

* Fine-tuning hyperparameters for optimal performance.
* Exploring additional regression models for project-specific requirements.