**Project Design Phase-I**

**Solution Architecture**

|  |  |
| --- | --- |
| Date | 06 May 2023 |
| Team ID | NM2023TMID15467 |
| Project Name | Project - Traffic Intelligence: Advanced Traffic Volume Estimation with Machine Learning |

**Solution Architecture:**

* Data Collection:The first step is to collect the relevant data for traffic volume estimation. This may include historical traffic volume data, weather data, road network information, and other relevant variables. Data can be sourced from traffic sensors, GPS devices, public datasets, or other data providers.
* Data Preprocessing: The collected data needs to be preprocessed to ensure its quality and compatibility with the Random Forest model. This may involve tasks such as data cleaning, handling missing values, outlier detection, and feature engineering. Feature engineering techniques can be applied to extract useful features from the raw data, such as time of day, day of the week, road characteristics, and weather conditions.
* Training Data Preparation: The preprocessed data is divided into training and testing datasets. The training dataset is used to train the Random Forest model. It is important to split the data randomly or based on temporal order to capture the temporal nature of traffic volume patterns accurately.
* Model Training: The training dataset is used to train the Random Forest model. The model learns the relationships between the input features and the target variable (traffic volume) during this stage. The Random Forest algorithm builds an ensemble of decision trees by randomly selecting subsets of features and data instances for each tree. The trees are trained to predict traffic volumes based on the selected features.
* Model Evaluation: After training, the performance of the Random Forest model is evaluated using the testing dataset. Various evaluation metrics can be used, such as mean absolute error (MAE), mean squared error (MSE), or R-squared, to assess the accuracy and generalization capability of the model. Model hyperparameters can be tuned to optimize performance.
* Model Deployment: Once the Random Forest model is trained and evaluated, it can be deployed for traffic volume estimation. The deployment can be done in various ways depending on the specific requirements and constraints of the application. For example, the model can be deployed on cloud infrastructure, edge devices, or integrated into existing traffic management systems. The deployment architecture should consider factors such as scalability, real-time performance, and data connectivity.

**Example - Solution Architecture Diagram:**



*Figure 1: Architecture and data flow of traffic volume estimation sample application*

**Reference:** [**https://www.sciencedirect.com/science/article/pii/S209575641530670X**](https://aws.amazon.com/blogs/industries/voice-applications-in-clinical-research-powered-by-ai-on-aws-part-1-architecture-and-design-considerations/)