**WEBSITE TRAFFIC ANALYSIS**



PHASE 2 SUBMISSION DOCUMENT

INTRODUCTION:

* Website traffic analysis is the process of examining and interpreting data related to visitors' interactions with a website. This analysis is crucial for understanding how users engage with a website, identifying areas for improvement, and making informed decisions to enhance the user experience, optimize conversions, and achieve specific goals.
* This project involves analyzing website traffic data to gain insights into user behavior, popular pages, and traffic sources.The goal is to help website owners enhance the user experience by understanding how visitors interact with the site.
* For achieving the goal, we can incorporate machine learning models to predict future traffic trends or user behavior patterns. This way of Incorporating machine learning models to predict future traffic trends or user behavior patterns in website traffic analysis can provide valuable insights for improving your website's performance and user experience.

There are few steps to be followed for incorporating the models in the website traffic as follows:

DATA COLLECTION AND PREPROCESSING:

* Gather historical website traffic data, including user interactions, page views, click-through rates, session duration, and any other relevant metrics through various sorces including web server logs, Google Analytics,analytics tools used by the website owner and datasets. For example we can include dataset from the **[www.kaggle.com](http://www.kaggle.com)**
* Ensure that your data is clean, well-structured, and includes a timestamp for each data point. This process of pre processing is required to handle issues like missing values, data consistency, and outliers and the data cleaning ensures that the analysis is based on accurate information.
* Normalize or standardize numerical features if necessary.
* Convert categorical variables into numerical format using techniques like one-hot encoding or label encoding.

FEATURE ENGINEERING:

* We can create relevant features that will help machine learning models better understand user behavior and traffic patterns. This could include features like day of the week, time of day, and seasonal trends.
* We can create lag features to capture time dependencies, such as the traffic in the previous hour or day.

DATA SPLITTING:

* We can split our dataset into training, validation, and test sets.
* The training set is used to train your machine learning models, the validation set helps tune hyperparameters, and the test set is for evaluating model performance.

MODEL SELECTION:

* We can choose appropriate machine learning models for the task. Time series forecasting models like ARIMA, exponential smoothing, or machine learning models like linear regression, decision trees, random forests, gradient boosting, or deep learning models can be suitable for predicting traffic trends.

MODEL TRAINING:

* We can train our selected models using the training data.
* We can also consider using cross-validation for hyperparameter tuning and to prevent overfitting.

EVALUATION METRICS:

* We can use appropriate evaluation metrics for the problem, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or custom metrics relevant to your website's objectives.

HYPERPARAMETER TUNING:

* We can tune the hyperparameters of your models using the validation set to optimize their performance.

MODEL EVALUATION:

* We can evaluate your models on the test set to assess their performance and make sure they generalize well to unseen data.

INTERPRETABILITY:

* For better understanding, we can use model interpretability techniques to explain why your models make certain predictions.

DEPLOYMENT:

* Once we satisfied with a model's performance, we can deploy it to a production environment. This can be integrated into website's backend or analytics tools to make real-time predictions.

CONTINUOUS MONITORING:

* Continuously monitor the model's performance in the production environment, and retrain it as needed to adapt to changing user behavior or traffic patterns.

VISUALIZATION:

* We can create visualizations and dashboards to present the predictions and insights to stakeholders and decision-makers.

FEEDBACK LOOP:

* We can collect feedback from users and stakeholders to further improve the model and the website's user experience.

PROJECT DOCUMENTATION DONE BY:

SAKTHIVEL.R (511321205026)

GIRIVASAN.K(511321205008)

POOVARASAN.M(511321205023)

SARAVANAN(511321205032)