**Web Traffic Time Series Forecasting**

**Problem**

Web traffic can be defined as the number of visits to a website, including requests sent and received by web users. We aim to predict future web traffic for approximately a total of 145k Wikipedia articles to make better traffic control decisions. The increase in traffic for the websites could cause a lot of inconvenience for the users by a crashed site or very slow loading time. Therefore, a traffic management technique or plan should be put in place to reduce the risk of such problems.

**Client**

The accurate forecasting of future web traffic will help website servers optimize traffic management effectively, and thus maintain their regular customer service level.

**Data**

The data is collected from the Kaggle competition and can be found at <https://www.kaggle.com/competitions/web-traffic-time-series-forecasting/overview>

The training dataset consists of approximately 145k time series. Each of these time series represent a number of daily views of a different Wikipedia article for a total of 803 days from July 1st, 2015 to September 10th, 2017. In the competition, the challenge was to build a model on it and predict future traffic on each of the page for 62 days from September 13th to November 13th, 2017.

Approach

* clean the dataset by imputing the missing data if any.
* separate the time series dataset into the groups that contain a certain language.
* explore the raw data of each group to check if there is any pattern on time.
* check stationarity by Augmented Dicky-Fuller (ADF) test.
* converting the data into stationary format by detrending, differencing, and transformation.
* build three different types of Time series models: ARIMA (Auto-Regressive Integrated Moving Average), Prophet Facebook, and LSTM (Long Short-Term Memory).
* compare the different models to find a better solution.
* forecast future web traffic by using the models and interpret the results.

**Deliverables**

* Jupyter notebook that contains exploratory data analysis, feature engineering, and modeling.
* Final report that includes the entire process from data cleaning to model solutions.
* Slide deck for presentation.