Daily Website Traffic Trends: A Time Series Analysis

Group 9 Team Members:

1. Priya Kumari (kumarip@iu.edu)
2. Ayesha Tajammul Ahmed Mulla (amulla@iu.edu)
3. Aazin Asif Shaikh (aazshaik@iu.edu)

Objective:

The project aims to create an application allowing analysis of any website traffic data to help identify patterns, trends, and seasonality in the data captured daily, and implement an interactive dashboard to forecast future website traffic trends based on historical data and make informed decisions patterns observed.

Project Description:

Background:

Web traffic is the amount of data sent and received by visitors to a website and it has been the largest portion of Internet traffic. Internet traffic flow prediction heavily depends on historical and real-time traffic data collected from various internet flow monitoring sources. With the widespread traditional traffic sensors and new emerging traffic sensor technologies, traffic data are exploding, and we have entered the era of big data internet traffic. Internet traffic management and control driven by big data is becoming a new trend. Although there have been already many internet traffic flow prediction systems and models, most of which use shallow traffic models and are still somewhat unsatisfying. This inspires us to reconsider the internet traffic flow prediction model based on deep architecture models with such rich amount of internet traffic data.

Usefulness:

Daily website traffic forecasting can be a useful tool for website owners and managers to predict future website traffic, plan marketing strategies, and optimize website performance. Some of the key benefits and applications of daily website traffic forecasting are:

i. Resource allocation: Daily website traffic forecasting helps website owners and managers to allocate resources effectively. They can determine how many servers, bandwidth, and other resources they will need to handle the expected traffic.

ii. Marketing planning: Daily website traffic forecasting helps website owners and managers to plan their marketing campaigns more effectively. They can predict the impact of their marketing efforts on website traffic and adjust their strategies accordingly.

iii. Performance optimization: Daily website traffic forecasting helps website owners and managers to identify potential performance issues in advance. They can optimize their website infrastructure, such as caching and load balancing, to ensure that their website can handle the expected traffic.

iv. Revenue forecasting: Daily website traffic forecasting helps website owners and managers to forecast revenue more accurately. They can estimate how much revenue they can expect to generate based on the predicted traffic levels.

v. Business growth: Daily website traffic forecasting can help website owners and managers to plan for business growth. They can identify trends in website traffic and use this information to expand their business.

Current Tools:

There are currently similar tools in the market such as Google Analytics, Flurry and Matomo which analyze web traffic data and give an overview about the trend and seasonality.

The major differences between the proposed application and the current tools in use are:

   i.   Customization: A custom dashboard application can be tailored to the specific needs and goals of a website owner or manager and can be customized to display the metrics and insights that are most relevant and useful to them. In contrast, Google Analytics provides a fixed set of metrics and features, although it can be customized but only to some extent.

   ii. Pricing: A custom dashboard application may involve upfront development costs, but there may be no ongoing fees or subscription costs. In contrast, while Google Analytics is free to use, some of its advanced features and integrations may require a paid subscription to Google Analytics 360.

Target Stakeholders:

Some targeted stakeholder groups for the proposed application would be as follows:

   i.   Marketing and advertising teams: They can use the dashboard to measure the effectiveness of their marketing campaigns, identify opportunities to optimize their campaigns, and track the ROI of their advertising spend.

   ii. IT and web development teams: They can use the dashboard to monitor website performance and uptime, identify and troubleshoot technical issues, and optimize website speed and user experience.

   iii. Executives and senior level-stakeholders: They can use the dashboard to gain a high-level overview of website traffic trends, understand the impact of website traffic on business goals, and make strategic decisions on resource allocation and investment.

Dataset:

Origin and Description:

The data contains 5 years of daily time series information for several measures of traffic on a course website. The variables exhibit complex seasonality that is represented using the day of the week and the calendar. The variables are daily counts of page loads, unique visitors, first-time visitors, and returning visitors to the website. Overall, there are 2167 rows of data spanning the date range from September 14, 2014, to August 19, 2020. A visit is defined as a stream of hits on one or more pages on the site on a given day by the same user, as identified by IP address. Multiple individuals with a shared IP address (e.g., in a computer lab) are considered a single user. A visit is classified as "unique" if a hit from the same IP address has not come within the last 6 hours. Returning visitors are identified by cookies if those are accepted. All others are classified as first-time visitors, so the count of unique visitors is the sum of the count of returning and first-time visitors by definition. Thus, this dataset is reliable and a good fit for time series analysis as it captures the daily web traffic and can be further analyzed for specific months, quarters etc.

Dataset URL: https://www.kaggle.com/datasets/bobnau/daily-website-visitors

Approach:

There are several approaches to forecasting website traffic. One common method is to use historical data to identify patterns and trends in traffic volume. This can involve analyzing data over long periods of time, such as several years, to identify seasonal or cyclical patterns. Alternatively, shorter timeframes, such as weeks or months, can be used to identify patterns that are more immediate.

Another approach is to use external factors that may influence website traffic, such as social media trends, weather patterns, and advertising campaigns. By analyzing these factors, website owners can make more accurate predictions about traffic volume and adjust their strategies accordingly. Since our data is of a course website, external factors would be semester intake, faculty assigned, students' major and class number and strength which influence the page loads, number of new visitors, reloads and other attributes.

GitHub URL: https://github.iu.edu/kumarip/TSA_amulla_kumarip_aazshaik.git

References:

1. https://www.researchgate.net/publication/349447175_Forecasting_Website_Traffic_Using_Prophet_Time_Series_Model