Web-Traffic Forecasting

Tejaswi Sai

2022-06-27

## Abstract

Web traffic forecasting in recent days has become a major problem as it can cause failures in the normal functionality of major websites. The application of machine learning algorithms for predictive models is a popular research topic. In the industry, where technology has been advancing daily, performing web traffic predictions has been a major problem. The process of web traffic forecasting has been experienced with several issues, including classification and analysis to forecasting. Therefore, to help solve these problems, a machine learning predictive model would be the best solution. This paper, therefore, has implemented a time series predictive model that helps in web traffic forecasting. The project has utilized a dataset already available on the Kaggle.com website. Apart from forecasting, the project has identified the factors for consideration in web traffic forecasting and their severity to the normal functionality of the major websites. RStudio has been used for this project.

## Introduction

With the increased availability of digital devices like phones and laptops, more and more people have access to the internet at any place and anytime all over the world, and this has caused an increase in traffic for most websites. The increased traffic poses a major problem to the websites and the hosting companies as they try to cope with the changes in traffic most effectively and efficiently. This is because they need to ensure no inconveniences are caused to the website users either by having a slow load time or crash site. Therefore, this brings in the need for the most suitable strategy to reduce such challenges. A solution to the challenges will be found with the development of a predictive model for web traffic forecasting in this paper.

##Literature Review

In network design and management, the prediction, analysis, and model simulation of the internet traffic always play a critical role. In the study by Chen et al. (2018), they propose a modified LSL algorithm that could predict the web traffic and modify parameters with high accuracy and fast convergence. The research applied a time series that ensured the adaptive ability of the predictive model (Chen et al., 2018). According to Kafi et al. (2014), it is always essential to have internet traffic information in a timely and accurate manner for applications like admission control, bandwidth allocation, congestion control, and anomaly detection. Therefore, with the identified previous works suggestions, there came a need to have this project where we implement the machine learning algorithm for web traffic forecasting. The project answers the research question of the main contributing factors to website traffic congestion. ## Data

## Rows: 144 Columns: 2

## ── Column specification ────────────────────────────────────────────────────────

## Delimiter: ","

## chr (1): date

## dbl (1): visits

##

## ℹ Use `spec()` to retrieve the full column specification for this data.

## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## # A tibble: 6 × 2

## date visits

## <chr> <dbl>

## 1 1949-01 112

## 2 1949-02 118

## 3 1949-03 132

## 4 1949-04 129

## 5 1949-05 121

## 6 1949-06 135

check for any missing values.

sapply(web\_traffic,function(x) sum(is.na(x)))

## date visits

## 0 0

Remove any duplicates.

unique\_data\_interactions<-unique(web\_traffic)

## Methodology

With the data having been cleaned next willinvolve analyis on the data and creating the predictoive model.

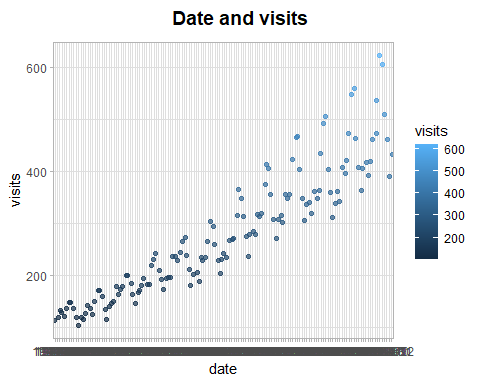
x <- ggplot(web\_traffic, aes(date,visits)) +

geom\_jitter(aes(color = visits), alpha = 0.7) +

theme\_light()

title <- ggdraw() + draw\_label("Date and visits", fontface='bold')

plot\_grid(title,x, ncol=1, rel\_heights=c(0.1, 1))

 ## Results

From the analysis, it is clear that there was an association between dates and the number of visits to the website by the users. In most cases, there was a high number of visits in days allocated for weekdays unlike weekdays. This is because, during weekdays alot of people are engaged in activities especially work related.

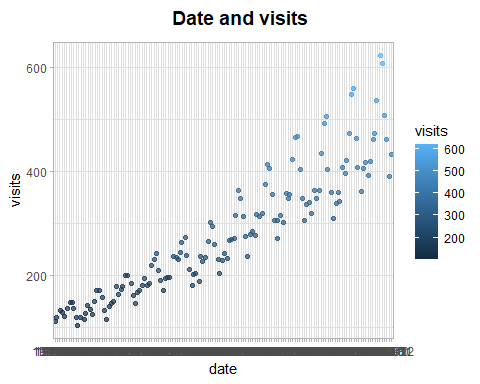
x <- ggplot(web\_traffic, aes(date,visits)) +

geom\_jitter(aes(color = visits), alpha = 0.7) +

theme\_light()

title <- ggdraw() + draw\_label("Date and visits", fontface='bold')

plot\_grid(title,x, ncol=1, rel\_heights=c(0.1, 1))



## Implications

For future works, I would be recommended that an NLP machine learning algorithm be implemented for web traffic forecasting. More sample data needs to be collected for proper and accurate analysis.

## Conclusion

The project performed an analysis of the web traffic and predictions on what factors affect the availability of web services. Time was the main factor since, at different times, there were varying visits to the website.

## References

Chen, J. F., Lo, S. K., & Do, Q. H. (2018). Forecasting short-term traffic flow by fuzzy wavelet neural network with parameters optimized by biogeography-based optimization algorithm. Computational intelligence and neuroscience, 2018.

Kafi, M. A., Djenouri, D., Ben-Othman, J., & Badache, N. (2014). Congestion control protocols in wireless sensor networks: A survey. IEEE communications surveys & tutorials, 16(3), 1369-1390.