



Web Traffic Analysis Using Machine Learning

Sindhu G¹, Kalarani R², Revathi K³, Megashree R⁴, Vishnu R⁵

^{1,2,3}Assistant Professor- Department of Computer Science and Engineering, Agni College of Technology, Chennai, India.

⁴Teaching Assistant- Department of Computer Science and Engineering, Agni College of Technology, Chennai, India.

⁵UG – Department of Computer Science and Engineering, Agni College of Technology, Chennai, India.

Email ID: sindhu.1498@gmail.com¹, kalamtechit@gmail.com², revathikannan31@gmail.com³, mega1791998@gmail.com⁴, vishnuramesh542@gmail.com⁵

Abstract

Web Traffic Analysis enhance the digital marketing strategies and user experience by analyzing website traffic, user engagement metrics, and customer behavior patterns. Time series analysis and forecasting in time series data finds it significance in many applications such as business, stock market and exchange, weather, medical, electricity demand, web traffic and user behavior, cost and usage of products such as fuels, electricity etc. Time series data contains a high volume of numerical data and a time dimension which captures a lot of information including inter-pattern trends and correlations. The Pattern Discovery Method (PDM) in Machine Learning is used for performing accurate analysis of time series data and making important decision. The Pattern Discovery Method which includes the process of pattern deploying and pattern evolving to improve the effectiveness of using and updating discovered patterns for finding relevant and forecasting information.

Keywords: Pattern Discovery, Time series data, Web Traffic, Correlation, Inter Pattern Trends.

1. Introduction

A time series is a succession of chronologically ordered data spaced at equal or unequal intervals. The forecasting process consists of predicting the future value of a time series, either by modelling the series solely based on its past behavior or by using other external variables. This document describes how to use machine learning in order to forecast the Web traffic. In Data Preparation phase the web log data must be cleaned, filtered, integrated and transformed in such a way that the irrelevant and redundant data can be removed, user session and transaction can identify. After data preparation phase, the pattern discovery method should be applied. This phase consists of different techniques derived from various fields such as statistics, machine learning, data mining, pattern recognition, etc. applied to the web domain and to the available data. The task for discovering the patterns offer some techniques as statistical analysis, association rules, sequential

pattern analysis, clustering and so on. [1-2]

2. Literature Review

"Optimizing compilers is a difficult and time-consuming task," especially when done by hand. As far as we know, the compiler handles both translation and optimization. An efficient compiler system can become more automated and simpler, as evidenced by recent studies using deep learning and machine learning approaches. Model training, prediction, optimization, and feature selection are handled by most machine learning and deep learning methods. In this case, choosing the optimal characteristics is necessary in order to use deep learning and machine learning techniques to enhance the optimization quality. "Web Traffic Prediction with Deep Learning: A Survey" This paper reviews the application of deep learning techniques in predicting web traffic, focusing on various models, such as recurrent neural networks (RNN) and long short-term memory

(LSTM). Time Series Forecasting with Machine Learning: A Survey" The paper focuses on machine learning methods for time series forecasting, particularly in areas like web traffic, and provides a comparative analysis of various models including ARIMA, SVM, and deep learning. Prediction of Web Traffic Using Machine Learning Models: A Comparative Study" This study compares various machine learning algorithms such as decision trees, support vector machines, and neural networks in predicting web traffic patterns. "Machine Learning Techniques for Web Traffic Analysis" The paper discusses the application of machine learning techniques such as clustering, regression, and classification in web traffic analysis, including the prediction of future traffic trends and user behavior. "Web Traffic Classification Using Machine Learning Algorithms" This research explores machine learning algorithms for classifying web traffic into different categories based on user interaction and browsing patterns.[5]

3. Attributes

Web traffic data does not just represent the number of people that visit the website. Other core attributes shown by the website traffic includes the length of time a given user stays on the website. This is assessed through metrics such as the time taken by a user on a page, unique visitors, average session duration, Pages per Session, traffic Sources, geographical data, and the bounce rate. If we have a business, web traffic data also shows the conversion rate which means it determines the rate of purchase by users. Furthermore, it also shows how much it costs to bring in a visitor to the website. These are examples of the specific metrics which can be used to determine the measure of web traffic.[6]

4. Web Traffic Measurement

Web traffic is determined by the popularity of websites and individual pages or parts within the site. We can do this by viewing the traffic statistics found in the web server log file, the automatically created list of all the pages served. The web traffic is measured based on unique visitors, average session duration, pages per Session, traffic Sources, geographical data, and the bounce rate. A hit is made when a file is recorded. Tracking out-of-site

applications will monitor traffic by adding a HTML code on each page of the website. Online advertisement can also be tracked easily using this data and to obtain attribution insights e.g. the number of clicks per user, numbers of visits on site, number of ads appearance, completion of an order by advertisement targeted users.[3-4]

5. Web Traffic Tools

Google Analytics is a web analytics platform introduced by Google for tracking and analyzing users' traffic to websites. Google Analytics is a widely used Analytics platform by businesses, that provides valuable detailed information about their online audience: performance matrix, integration capability, and powerful features. Businesses can get detailed information on website visitors, which includes their interests, location, age, gender, and device used to access websites such as mobile, desktop, or tablet. Google Analytics helps businesses to track the traffic source of their websites which includes organic search, traffic gained using referrals, social media, and traffic gained through paid advertisement campaigns. Adobe Analytics is an extensive web analytics platform designed to help businesses understand and optimize users' online presence. It is offered by Adobe Inc., Adobe Analytics helps businesses track, analyses, and interpret visitor's actions on websites. Hotjar Web analytic platform offers wide a range of features that help businesses to understand website user's behavior and optimize their experience. Clicky, which offers end-to-end solutions for tracking and analysis that help businesses to understand their visitor's behavior and interest. SEMrush is a digital platform that comes up with a large range of tools and features that help businesses improve their user experience, and online visibility of the business, and improve visitor traffic. With their keyword research tool businesses can identify phrases that are relevant to their business, service, and product, this helps businesses to rank their website at the top when users search related keywords in search engines.[7-10]

6. Problem Statement

Existing web traffic tools are expensive and primarily designed for business applications. The proposed system calculates web traffic based on key metrics

and provides alternative solutions based on user requests. Initially, the web log files are cleared and used to collect and store user input. The web server log is maintained to track and display web traffic data, such as the number of page visits, visit duration, and bounce rate. A pattern discovery method is then applied to provide solutions and recommend web pages based on user requests. (Figure 1)

6.1. Architecture Diagram

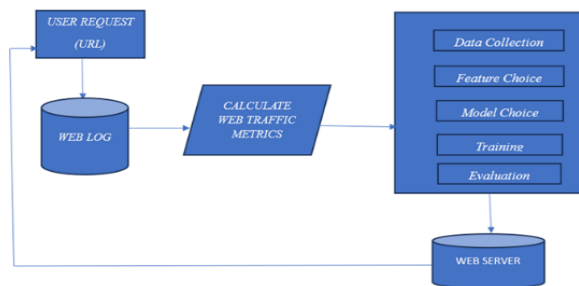


Figure 1 Architecture Diagram

6.2. Methodology

Problem Statement: Analysis of Web traffic Using Pattern Discovery method

Input: URL from User

Output: To display Web traffic

Step:1 Creation of Web log

Step:2 Traffic Analysis Metrics and Formulas

2.1 Unique Visitors (UV)

Formula: Unique Visitors = Total visitors – Returning Visit

2.2 page views(pv)

Formula: sum (pages viewed by each visitor)

2.3Pages Per Session (PPS)

Pages Per Session =Total page views/total sessions

2.4 Average session Duration

Average session Duration= Total Time Spend by All users/total sessions

2.5 Bounce Rate (BR)

Bounce Rate =(single page session / total sessions) X 100

2.6 Traffic Sources Breakdown

Traffic Sources percentage=(visitors from a specific source / Total Visitors) X 100

Step:3 Applying Pattern Discovery method

3.1 Data Collection

3.2 Feature Choice

3.3 Model Choice

3.4 Training

3.5 Evaluation

Step:4 Web Server

4.1: Display web traffic metrics

4.2: "Display recommended web resources based on specific metrics, and adapt the content in response to user-requested modifications."(Figure 2)

6.3. Implementation

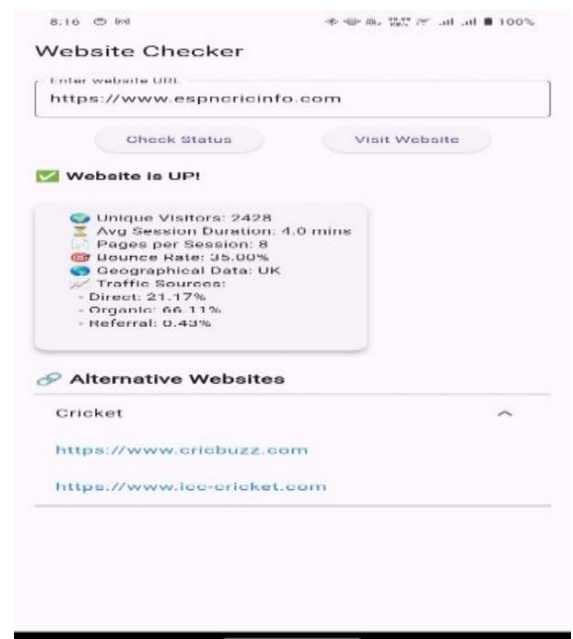


Figure 2 Implementation

Conclusion

In conclusion, Web Traffic Analysis plays a crucial role in enhancing digital marketing strategies and user experience by providing deep insights into website traffic, user engagement, and behavior patterns. Time series analysis offers significant advantages in forecasting and understanding trends across various domains, including business, weather, and web traffic. The use of advanced techniques like the Pattern Discovery Method (PDM) in Machine Learning further improves the accuracy and effectiveness of time series data analysis. By evolving and deploying patterns, PDM ensures continuous improvements in decision-making and forecasting, thus enabling businesses to adapt to



changing dynamics and optimize their strategies accordingly.

References

- [1]. Jan Luxemburk;Karel Hynek ,Towards Reusable Models in Traffic Classification, 2024 8th Network Traffic Measurement and Analysis Conference (TMA)
- [2]. Lal Bahadur Pandey;Manisha Sharma;Rajesh Tiwari;Radhe Shyam Panda;Partha Roy,A Survey of Optimized Compiler Using Advanced Machine learning and Deep Learning Techniques, 2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Application (ICCCMLA)
- [3]. William; Gurpreet;Abhinandan Kondekar;VishalM.Tidake;Shrikaant Kulkarni;P. B. Narendra Kiran ,Edge based Web Computing for Traffic Control Management for Distributed Environment Conditions, 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)
- [4]. Adwitiya Sinha;Ratik Puri;Udit Balyan;Ritik Gupta;Ayush Verma ,Sustainable Time Series Model for Vehicular Traffic Trends Prediction in Metropolitan Network, 2020 6th International Conference on Signal Processing and Communication (ICSC)
- [5]. Smyl, S. "Time Series Forecasting with Machine Learning: A Survey", International Journal of Data Science and Analytics, 2020.
- [6]. Thomas Rincy N;Gupta, A Survey on Machine Learning Approaches and Its Techniques: 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)
- [7]. Kumar, A., & Yadav, R. "Analyzing Web Traffic and Predicting User Behavior with Machine Learning Algorithms" Journal of Computer Science, 2019
- [8]. Venkatesh, G., & Gokulakrishnan, V. "Prediction of Web Traffic Using Machine Learning Models: A Comparative Study" IEEE International Conference on Machine Learning and Applications (ICMLA), 2019.
- [9]. Chien, S., & Chang, J., "Machine Learning Techniques for Web Traffic Analysis" International Journal of Artificial Intelligence, 2018.
- [10]. Shanmuganathan, V., & Rajendran "Web Traffic Prediction with Deep Learning: A Survey,International Journal of Computer Applications, 2017.