Project Proposal

On

VISUALISE WEBSITE CLIENTSIDE DATA

for

Certification on Big Data

By

**ManshiAgrawal(171500182)**

**Nancy Varshney(171500203)**

**Neha Chaudhary(171500205)**

**Shreshth Bharadwaj(171500328)**

Under the Supervision of

**Mr. Vaibhav Diwan**



Department of Computer Engineering and Applications

IET, GLA University-Mathura

**Contents:**

1.Introduction

2.Clickstream Data

3. Potential Uses of Clickstream Data

4. Flow chart to achieve objectives

5.Conclusion

**INTRODUCTION**

In today’s date, using website for getting any information is very common. We all use websites for someinformation and in doing so, lots of data are generated which are defined by ourclicks. Theseclicks are stored as log files in the valueof timestamp and the URL of the website which user havebeen visiting. Clickstreams are also used as a way to analyse the behaviour of the user. Let us take an exampleof an E-commerce website: User log into the website and then browse through theitems he/she wants tobuy from it. User adds theitem to thecart and goes on tobrowse some moreitems tobuy or maybe view only. On every view useris making, the user must beclicking theitem. That particularclick is logged as a clickstream log data and it can be used toread customer behaviour . Now the most common question arrives is how a clickstream log data can help in analysing the behaviour and thinking of a customer. The answeris hidden in big data analytics. There are sequential mining algorithms that are used for analysing customer behaviour and customerreadings. Clickstream data can demonstrate what terms an individual connected into a webbrowser torequest to access the desired page. These points of interest can uncover someother sites theclient may have goneby before accessing thecompany's site. Once the user has achieved the website, clickstream information can indicate what page theindividual initially arrived on, what highlights orcapacities heor she tapped on once there, how much time was spent on that page and where theonlineexcursion took him or her after that page. In this project Hadoop is used toextract information from clickstream data and useit to decrease theevents of cart abandonment. Hadoop helps toextract, store and analyse theclickstream or web log data and mergeit with the traditional customer data - in order to get betterinsights into the behaviour of the visitor and optimize the path to product buying.

**CLICKSTREAM DATA**

Clickstream data is an information trail a user leaves behind while visiting a website. It is typically captured in semi-structured website log files.

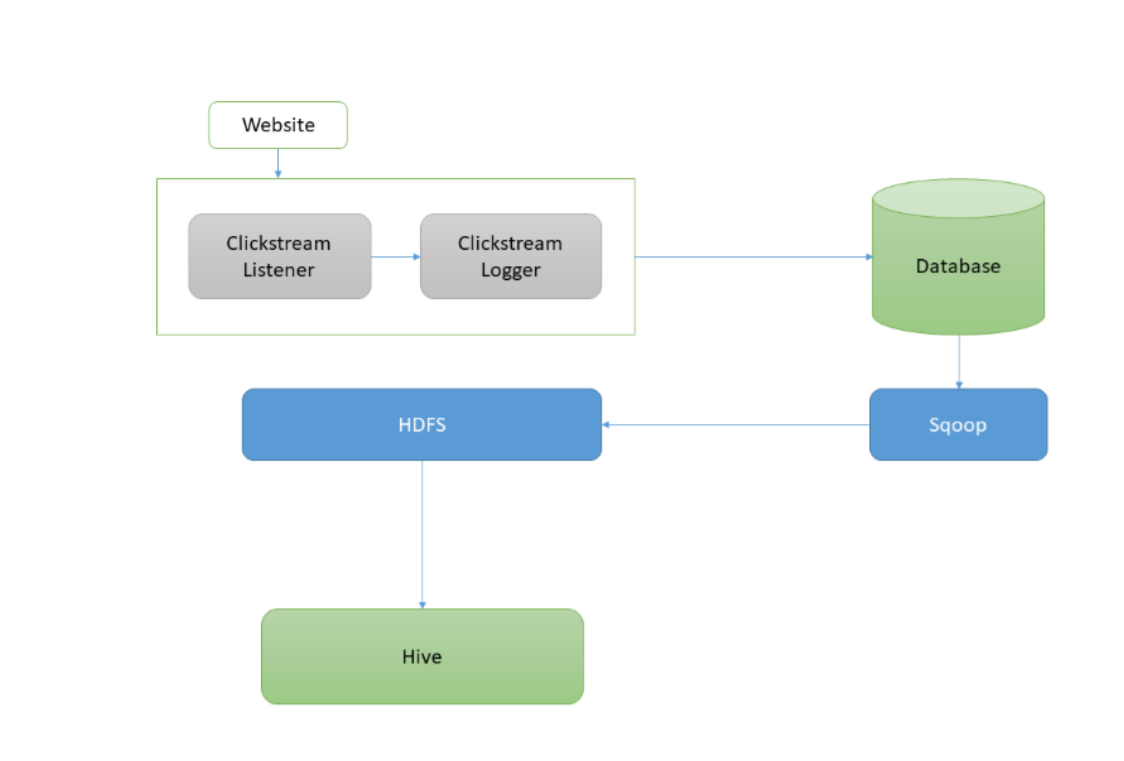
These website log files contain data elements such as a date and time stamp, the visitor’s IP address, the URLs of the pages visited, and a user ID that uniquely identifies the user.

**POTENTIAL USES OF CLICKSTREAM DATA**

One of the original uses of Hadoop at Yahoo was to store and process their massive volume of clickstream data. Now enterprises of all types can use Hadoop and the Hortonworks Data Platform (HDP) to refine and analyze clickstream data. They can then answer business questions such as:

* What is the most efficient path for a site visitor to research a product, and then buy it?
* What products do visitors tend to buy together, and what are they most likely to buy in the future?
* Where should I spend resources on fixing or enhancing the user experience on my website?

**FLOWCHART TO ACHIEVE OBJECTIVES**



**CONCLUSION**

From the analysis, it can beconcluded that clickstream data helps in many ways for a valuable profit to theorganization. Organizations can gain valuable insight to apply it to their 0% 50% 100% Accuracy Mean Absolute Error Relative Absolute Error Pattern Pruning Time Series Neural Network C5.0 International Journal of Pure and Applied Mathematics Special Issue 3731 websites forreading customer behaviour and then giving the priority of what the user wants and how. Thebest analysis till now to visualize data is shown tobe data pruning which is also known as pruned decision tree algorithm. The analysis shows about the use ofclickstream data and how valuableinformation insights from the log data of a clickstream can be useful to any organization.