International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V70I4P101

Volume 70 Issue 4, 1-12, April 2022

Original Article

Development of a Process-Driven Model for Extracting User Behavioral Pattern from Web Access Log Files

Folasade Adedeji[1], Adebowale Adekunle[2], Adewale Adebayo[3], Olujimi Alao[4]

[1,2,3,4] *Department of Computer Science, Babcock University, Ogun State, Nigeria.*

Received: 03 March 2022
Revised: 19 April 2022
Accepted: 23 April 2022

Abstract *- Investigating user behaviour with a focus on the process by which users achieve their aims has the potential of discovering additional insight about their actions, improving the process advertently improves their use of the website. This study aims to develop a Process-Driven model that extracts the information contained within web server logs to identify the patterns hidden in user behaviour to improve website usability. Data for this study was collected from the ICT unit in southwest Nigeria. The data contained 879475 records of web server logs of users. The data was preprocessed, cleaned, and transformed from a spreadsheet format into an event data format (.xes). Process mining was applied to the dataset using five (5) algorithms, following which their performances were compared based on simplicity, fitness, precision, and generalization. The results showed that the most optimized process model was developed using an Inductive miner. A Petri net model was used to define the users' navigational behaviour on the website. The study concluded that information extracted from these patterns could improve website design, thus impacting website usability.*

Keywords *- User Experience, Process mining, Web mining, Website usability, User behaviour.*

I. INTRODUCTION

Usability is how specific users can use a system, product, or service to achieve intended goals with effectiveness, efficiency, and satisfaction within a context of use [1]. Usability Engineering (UE) in the field of study is concerned with the way humans interact with the computer; it focuses on optimizing users' performance through standards, principles, and evaluations to produce highly usable interfaces [2]. UE involves a good understanding of the user and the interface or system to build a better system or computer interface. The end product is to have a user interface that makes users' experience as smooth and efficient as possible.

User Experience (UX) emerged as a wider and more complex approach compared to usability. In addition to usability, UX involves emotional and cognitive areas of human experiences, such as aesthetics, hedonism, and context [3].

Researchers agree that usability is one of the important components of UX [4]. Usability evaluation is done for several products and systems to locate problems with the interface design or its usage.

One of the areas of application is web applications, specifically websites. Recently, the use and dependence on websites have grown rapidly [5]

But not all users find it easy to use websites because of usability issues such as poor navigation, inconsistency, poorly structured links, and lack of relevant information. Bad usability is a major challenge that leads to low page hits and website failure [6]. Website owners and designers have to maintain a usable website to stay ahead of the competition. Analyzing user behaviour gives information about different characteristics of the users being studied; this helps to find areas to focus on [7]. Website usability evaluation studies have combined data capture and analysis methods into several models [8].

Investigating user behaviour with a focus on the process by which users achieve their aims has the potential of discovering additional insight about their actions, improving the process advertently improves their use of the website [9] [24]. Research shows increasing dependence on websites to carry out our day-to-day activities. Social distancing has made most businesses want to have a presence on the Internet. The use of web applications is almost endless, and they are beneficial as they improve productivity and streamline processes. The successful development of a process-oriented usability evaluation method is beneficial for improving the usability of websites and enhancing user experience.

A process is a collection of individual activities or events performed to achieve a certain goal. Process Mining (PM) is a method of discovering insight and gaining knowledge from processes by analyzing event data obtained during the execution of the process. This is done to discover a graphical model required for improving the underlying process [10]. Process Mining is an emerging field of study, and it extends data mining by adding the process perspective [11]. It stands as the bridge between process science and data science [12].

As a result, the study developed a process mining model that shows the website owners and designers a first-hand insight into what users are doing on their websites. Information gathered from the analysis of user behaviour was used to predict usage patterns and give recommendations on improving user processes on the website.

II. LITERATURE REVIEW

A number of works have been reviewed within the subject matter of analyzing user behaviour on various websites. Following the review, a number of research gaps were identified and presented as follows.

Existing literature applied descriptive statistical analysis of weblog files presented using tables and failed to present information using graphical plots, which are visual and reflect the underlying behaviour of users visiting websites [25].

Existing literature focused on the implicit analysis of user behaviour on the website by focusing on quantitative analysis of data rather than on explicit analysis of user behaviour which is qualitative, thereby guiding website developers on the navigational patterns of users on their websites.

Existing literature failed to implement the recommendations for the redesign of the website based on the analysis conducted, thereby re-evaluating the impact of the recommendation on website usability.

As a result of the gaps identified in the related works, this study focused on applying process mining to the analysis of web server logs collected from the website of an academic institution. The study used this approach to perform a process discovery of user navigational behaviour of the website with the need to recommend necessary modifications to the website that are necessary for improving website usability by the users.

A. Related Works.

[13] Assessed the user behaviour of a library website using information collected from web access logs. The study aimed to use behavioral models and web analytics to improve the usability and functionality of a library website in Russia. The study collected secondary data containing user interaction with the website from Google Analytics and Yandex. Metrica reports. The results of the web analysis showed the graphical plot of users' access from the login page to other pages within the website. The results of the web analysis were used to optimize the page structure of the library website by introducing a recommendation mechanism. The study failed to conduct a usability analysis following the adjustments made to the library website.

[14] Studied the usability of Library websites of Universities in the South-South region of Nigeria. 11 Library websites were evaluated; data collection was done by administering a survey. A checklist with 5 attributes usefulness, efficiency, effectiveness, learnability, and accessibility, was adopted. The result of the analysis showed the usability scores of each university based on the usability metrics employed. The authors concluded that university library websites need improvement in many usability aspects.

[7], worked on the classification of user behaviour using Neuro-Fuzzy logic modeling. The study collected a dataset composed of users' temporal logs such as local machine, network, and web usage logs following the collection of the users' 360-degree feedback. Various rules were implemented to address the company's policy for determining the precise behaviour of users. User behavior classification was achieved using a Gaussian Radial Basis Function Neural Network (GBRF-NN) trained on the sample dataset generated by a Rule-based Fuzzy System (FRBS) and the 360-degree feedback of users.

[15], devised a usability metric for evaluating University websites to see if they meet the needs of students. University websites in Canada, Europe, and the United States were evaluated. The evaluation was based on the content, navigation, design layout, organization, communication, and ease of use. A survey designed based on the metrics was conducted and completed by 265 students. The results showed that most university websites are usable; however, the usability standard desired by most students was not met. The study concluded that the metrics devised can be used to increase enrolment from potential students visiting the websites by improving usability.

[16] worked on the analysis of user behaviour on an e-commerce website. The study collected data consisting of information about user behaviour that was stored in web server logs of an e-commerce website. The study applied a linear-temporal logical model to analyze the weblog files. The results showed that event logs were generated by analyzing the weblog files. The event logs were used to capture the behaviour of users on the e-commerce website. The study provided recommendations for improving the website design due to the analysis to improve website

efficiency by users. The study did not consider the implementation of recommendations on a website redesign, thereby justifying its impact on improving user navigation of the e-commerce website.

III. MATERIALS AND METHOD

This section contains the procedures undertaken in this study to apply process mining to the web server logs collected from a university website. A systematic literature review technique known as the Preferred Reporting for Systematic Reviews and Meta-Analysis (PRISMA) approach was used to analyze the collected literature to identify existing website usability evaluation methods and related problems [17]. Data containing information that is relevant to user behaviour was collected from web server logs of a higher academic institution in south-western Nigeria and stored as a comma-separated variable (.csv) file format.

The collected web log files will be preprocessed by performing data cleaning. The weblog files were converted into an event log file stored in the extensible event stream (.xes) file format. The event log file generated will be subjected to process discovery analysis using process mining algorithms to generate the process model of the navigational behaviour of users as a tree plot and Petri net graph. The process model generated will be evaluated based on performance evaluation metrics such as simplicity, precision, fitness, and generalization and used as a basis for selecting the best model.

A. Research Framework

This study adopted the use of the process-driven model for website usability evaluation because it has the potential to show more insight into user behaviour by analyzing the process that users go through while browsing websites. Figure 1 shows a diagram of the research framework adopted in this study. As shown in Figure 1 below, the framework has 4 stages: Data capture, collection and preprocessing stage, data analysis stage, process discovery stage, and recommendation and enhancement stage.

B. Data Capture and Transformation

During the data capturing stage, event data of users' website activities were collected and extracted from the log files of the server used to host the website. Figure 2 shows a sample of a single log entry for a user accessing information from a resource made available on the website.

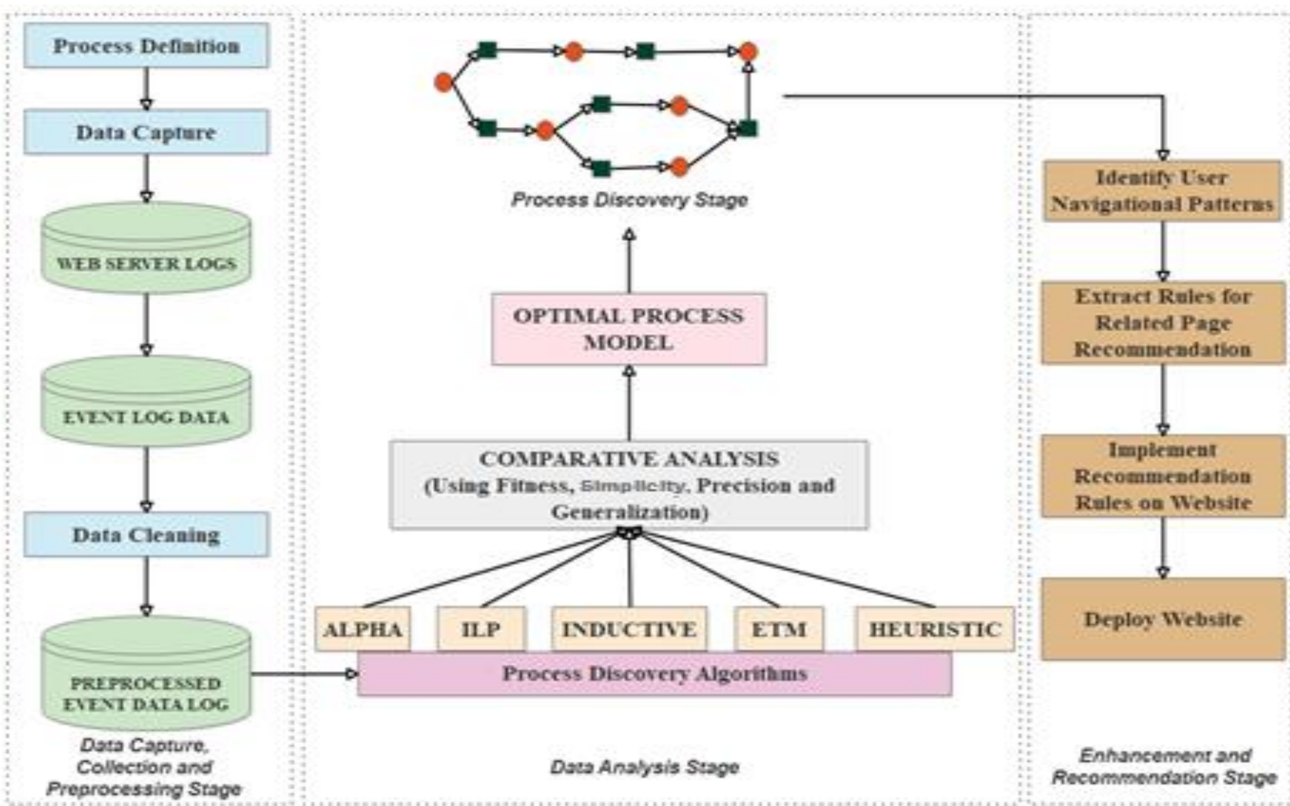


Fig. 1 Process-Driven Website Usability Evaluation Model (PDWUEM)

```
xxx.xxx.xx.xxx - - [14/Sep/2019:13:11:37 +0000] "GET
/alumni/books.html HTTP/1.1" 200 -"-" "Mozilla/5.0 (iPhone;
CPU iPhone OS 10_3_3 like Mac OS X) AppleWebKit/603.1.30
(KHTML, like Gecko) GSA/62.1.220348572 Mobile/14G60
Safari/602.1"
```

Fig. 2 Single web server log entry

The IP address represented with xxx.xxx.xx.xxx is a unique identifier of a user, represented by 4 octets (numbers between 0 and 255), followed by the date and time of accessing the resource, which is referred to as timestamp. The timestamp indicated by the values 14/Sep/2019:13:11:37 shows that the user accessed the information on the 14th of September, 2019, at 1:11 PM. Following the timestamp is the type of request made, which is indicated by a GET statement followed by the URL of the resource that has been accessed from the website. This request is the activity performed by a user accessing a webpage within the alumni page of the website as indicated by the alumni/book.html address. The request is followed by the web browser (Mozilla v5.0), the computing device (iPhone), and the operating system (iOS) that has been used to access the resource by the user. Figure 3 shows the webserver log dataset downloaded as a comma-separated variable (.csv) file.

1	ip-address	timestamp	web content accessed (GET) and actions (POST)	status	bytes	Resource
2	141.101.76.97	28-Feb-2021:12:00:30	GET /contactus HTTP/1.1	301	239	Mozilla/5.0 (Linux; Android 8.1.0; TECNO CF7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.110 Mobile Safa
3	141.101.105.39	28-Feb-2021:12:00:30	GET /contactus HTTP/1.1	200	18226	Mozilla/5.0 (Linux; Android 8.1.0; TECNO CF7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.110 Mobile Safa
4	172.69.54.24	28-Feb-2021:12:00:36	GET / HTTP/1.1	200	16756	Mozilla/5.0 (iPhone; CPU iPhone OS 14_2 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) CriOS/87.0.4280.77 Mol
5	141.101.69.34	28-Feb-2021:12:01:06	GET / HTTP/1.1	200	16756	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4324.190 Safari/537.36
6	141.101.69.86	28-Feb-2021:12:01:22	GET /university-e-learning-using-moodle HTTP/1.1	200	22308	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/14.0 Safari/605.1.15
7	173.245.54.60	28-Feb-2021:12:01:27	GET / HTTP/1.1	200	16756	Mozilla/5.0 (Linux; Android 4.2.1; en-us; Nexus 5 Build/JOP40D) AppleWebKit/535.19 (KHTML, like Gecko; googleweblight) Ch
8	141.101.104.80	28-Feb-2021:12:01:28	GET / HTTP/1.1	200	16756	Mozilla/5.0 (Windows NT 6.1; Win64; x64; rv:86.0) Gecko/20100101 Firefox/86.0
9	162.158.75.117	28-Feb-2021:12:01:43	GET /content/pg-school HTTP/1.1	200	13977	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.0; Trident/5.0; Trident/5.0)
10	172.69.55.187	28-Feb-2021:12:02:06	GET /university-e-learning-using-moodle HTTP/1.1	200	22308	Mozilla/5.0 (Linux; Android 10; TECNO KE5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.99 Mobile Safari/5:
11	172.69.55.187	28-Feb-2021:12:02:21	GET /university-e-learning-using-moodle HTTP/1.1	200	22308	Mozilla/5.0 (Linux; Android 10; TECNO KE5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.99 Mobile Safari/5:
12	172.69.253.144	28-Feb-2021:12:02:33	GET / HTTP/1.1	200	16756	Mozilla/5.0 (Linux; Android 8.1.0; TECNO LA7 Pro) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/84.0.4147.125 Mobile S
13	172.69.55.126	28-Feb-2021:12:02:50	GET /availablecourses HTTP/1.1	200	14804	Opera/9.80 (Android; Opera Mini/7.5.35721/191.218; U; en) Presto/2.12.423 Version/12.16
14	172.69.253.139	28-Feb-2021:12:02:52	GET / HTTP/1.1	301	230	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4324.150 Safari/537.36
15	172.69.253.145	28-Feb-2021:12:02:53	GET / HTTP/1.1	200	16756	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4324.150 Safari/537.36
16	141.101.105.13	28-Feb-2021:12:03:00	GET / HTTP/1.1	200	16756	Mozilla/5.0 (Linux; U; Android 7.0; TECNO P701 Build/NRD90M; wv) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Ch
17	141.101.105.13	28-Feb-2021:12:03:30	GET / HTTP/1.1	200	16756	Opera/9.80 (SpreadTrum; Opera Mini/4.4.33961/191.218; U; en) Presto/2.12.423 Version/12.16
18	162.158.94.81	28-Feb-2021:12:03:39	GET /content/faculty-education-1 HTTP/1.1	200	34519	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:82.0) Gecko/20100101 Firefox/82.0
19	162.158.89.40	28-Feb-2021:12:03:43	GET /content/20122013-post-utme-exercise HTTP/1.1	200	199708	Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:84.0) Gecko/20100101 Firefox/84.0
20	162.158.89.88	28-Feb-2021:12:03:51	GET /content/20122013-post-utme-exercise HTTP/1.1	200	199708	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_6) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/13.1.2 Safari/605.1.15
21	162.158.88.215	28-Feb-2021:12:03:53	GET /content/faculty-education-1 HTTP/1.1	200	34519	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/14.0.1 Safari/605.1.15
22	172.69.63.142	28-Feb-2021:12:04:03	GET /profiles/Netframe HTTP/1.1	301	247	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
23	172.69.63.206	28-Feb-2021:12:04:03	GET /service HTTP/1.1	301	237	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
24	162.158.79.248	28-Feb-2021:12:04:03	GET /bbva HTTP/1.1	301	234	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
25	173.245.54.162	28-Feb-2021:12:04:03	GET /My HTTP/1.1	301	232	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
26	173.245.54.228	28-Feb-2021:12:04:03	GET /modules/dashboard HTTP/1.1	301	247	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
27	162.158.79.34	28-Feb-2021:12:04:03	GET /wp-content/themes/twentyfifteen HTTP/1.1	301	261	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
28	173.245.54.174	28-Feb-2021:12:04:03	GET /wp-content HTTP/1.1	301	240	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
29	162.158.79.166	28-Feb-2021:12:04:03	GET /manage HTTP/1.1	301	236	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53
30	162.158.79.248	28-Feb-2021:12:04:03	GET /themes HTTP/1.1	301	236	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/53

Fig. 3 Web server logs downloaded as comma-separated variable (.csv) file

The downloaded web server log file contained 879475 records of user activities which were described based on information about six (6) attributes. The six attributes identified include the IP address of the computing device used, timestamp, web content accessed (GET) and actions taken (POST), status response, number of bytes used, and resources.

At the transformation stage, the collected web server log file was transformed by performing a number of activities required for presenting the dataset as an event log data presented as an extensible event stream (.xes) file. The data transformation will be performed using a number of scripts implemented using the python programming language. The pm4py, a process mining package, was installed on the python anaconda IDE to convert the web server log file into an event log file. Figure 4 shows a screenshot of the python script used to import the CSV file and convert it into an event log data stored as an. xes file.

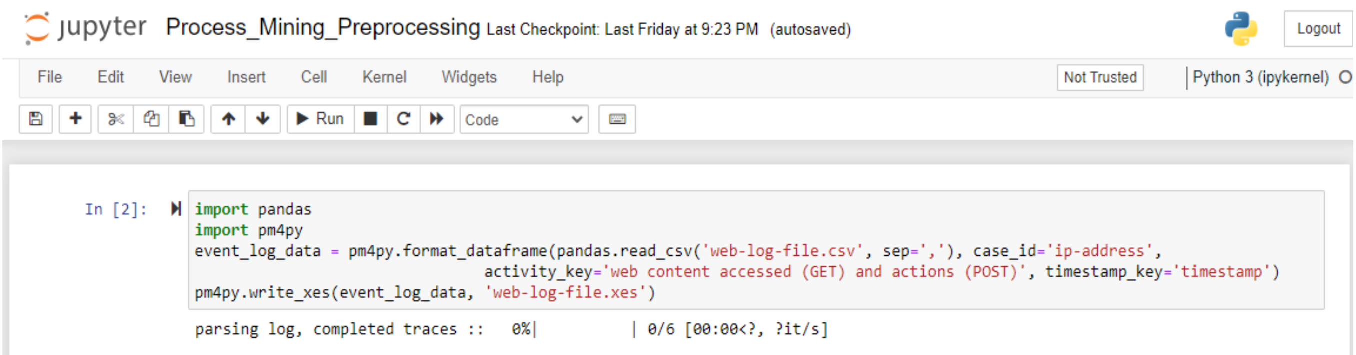


Fig. 4 Screenshot of conversion of .csv file into .xes file

The pm4py package was applied to the CSV file for conversion into the xes file after the panda package was used to import the CSV file into the python Jupyter notebook. As indicated in the figure, the first two lines are used to import the panda's package (line 1) and import the pm4py package. Line 3 will be required to import the CSV file (called web-log-file.csv) using the pandas and convert the file into an event log data stored in xes format. During the conversion, the IP address will be tagged as the case_Id, the web content accessed (GET), and actions (POST) will be tagged as the activity_key, while the timestamp will be tagged as the timestamp_key. The last line will store the event log file as a .xes file titled web-log-file. xes in the system directory. The eXtensible Event Stream is the standard format supported by the IEEE task force on process mining. This standard provides a generally XML format for exchanging event data between information systems in several domains where applicable and analysis tools for such data.

C. Data Cleaning and Preprocessing

At this stage, the data were subjected to cleaning and other preprocessing techniques to remove content that may affect the intended purpose of the data. Event The log data was cleaned by removing references to graphics, bot actions, sound files, and style entries contained within the entries of the records of the event log data. The event log was rid of rows with blank or missing columns values. Once the event log had been obtained, it was imported into the ProM 6 Framework and converted to the eXtensible Event Stream (XES) format. The eXtensible Event Stream is the standard format supported by the IEEE task force on process mining.

D. Data Analysis

At this stage, the event data was subjected to analysis by applying a number of process mining algorithms to extract the patterns hidden within the event log data. The algorithms were applied one after the other to the event log data, following which the performance of the algorithms was evaluated based on a number of performance evaluation metrics. The data analysis was performed using the ProM tool software, Java-based software, to perform process mining techniques on event log data. The process model was presented on the ProM tool as a Petri net model, which was used to visualize the navigational patterns of the users of the website based on the information extracted from web server logs. In this study, five process mining algorithms were used to extract the patterns existing within the event log data from the webserver log. A description of each algorithm is presented as follows:

a) Alpha Miner

The alpha miner is the first type of process mining algorithm aimed at examining casualty from an event log data. The goal is to generate a workflow net based on relations observed between activities in an event log data.

b) Heuristic Miner

This algorithm derives a set of activities connected using XOR and AND connectors to and from dependent relations (activities). It is based on a heuristic approach, making it practical for use in process discovery. It is highly recommended on event data that does not have too many unique events and usually produces a heuristic net.

c) Fuzzy Miner

This algorithm is useful for large numbers of activities and highly unstructured behaviour. It can be used to animate the event log on top of a created model to understand the dynamic process of system behaviour. It is recommended for use when complex and unstructured log data is available or when simplicity is required.

d) Inductive Miner

Inductive miner relies on building a directly following graph from the event log and using this graph to detect various process relations. It relies on detecting various cuts on the directly following graph created using the event log. The inductive miner lies in the unique methodology of discovering various divisions of the arcs in the directly-follows graph and using the smaller components after division to represent the execution sequence of the activities.

e) Integer Linear Programming (ILP) Miner

The Integer Linear Programming (ILP) miner gives a partial ordering over all regions of a language that allows for searching unrelated and minimal regions. This way, the unnecessary addition of places that have little added value to the behavioral restriction of the net is avoided. The ILP mining technique focuses on finding a Petri net that very precisely represents the behaviour as it is seen in the log.

E. Performance Evaluation Metrics

This study considered the optimal algorithms from the result of four-performance metrics. The overall fitness of the algorithms will be calculated by adding up the values of each of the quality dimensions' precision, generalization, fitness, and simplicity. The values range from 0 (worst) to 1 (best). The overall fitness of the algorithms is defined as:

a) Fitness

The process model retrieved via process mining techniques should represent the behaviour of the log to give insight into the underlying process [21]

b) Simplicity

The process model retrieved via process mining techniques should focus on the important aspects of the underlying process instead of showing all details in the log at once [21]

c) Precision

This metric shows the number of behaviour that is presented by the model but are not present in the event log [23]

d) Generalization

Generalization assesses the extent to which the resulting model will be able to reproduce future behaviour of the process.

The four metrics were computed on a scale from 0 to 1, where 1 is optimal. There is the possibility of the three metrics, fitness, simplicity, and precision getting up to optimum value. But in the case of generalization, it can have the optimum value of 1 in the limit, i.e., the value gets closer to 1 in relation to the frequency with which the node is visited.

IV. RESULTS

This section presents the results of the methods adopted in this study to develop the process mining model required for defining users' navigational behaviour on a website. The analysis of the existing methods of website usability evaluation was done using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting standard was adopted in selecting the relevant literature. PRISMA was developed by [18].

The results showed that one hundred and twenty-two (122) papers were identified from databases such as Google Scholar, Science Direct, and IEEE. Search parameters such as "Evaluation of Academic website," "Usability of academic websites," Usability of higher education websites," and "University website evaluation" were used to search for articles. A total of one hundred and twenty-two (122) works of literature are in the identification stage. After removing duplicate copies, eighty-four (84) were included in the screening stage.

Usability evaluation for visually impaired mobile interfaces was excluded because the focus of the study is mainly on websites. Titles that included usability, quality, accessibility, and navigation were included. In the eligibility stage, the full texts of the articles were studied, focusing on their relevance. At this stage, different metrics, elements, and methods used by researchers were identified, as well as the usability problems found. Fifty-two (52) kinds of literature were not relevant to the study. A total of thirty-two (32) were in the inclusion stage. Figure 5 shows the result at each stage of applying the PRISMA approach adopted in this study.

Table 1 shows the categories of the common usability issues that were identified from the assessment using the PRISM technique. Various issues that were related to navigation from existing methods of evaluation include: broken links, pages not found, and inappropriate labeling, to mention a few. This led to studying the navigation behaviour using the process-driven model to identify the unforeseen insight that exists within activities taking place on the website.

Table 1. Usability problems identified in the literature

Usability Issues	
Interface design	• Missing Alternate text • Errors in code (e.g., html errors) • Design outdated • Empty labels • Not mobile-friendly • Lack of icon description • The search box is too small
Navigation	• Broken links • Page not found • Inappropriate labeling
Content	• Instant feedback missing • FAQ not available • Missing information • Search returning empty information
Performance	• Slow loading time • Lack of security due to outdated app
Accessibility	• Non-conformance to WCAG standards

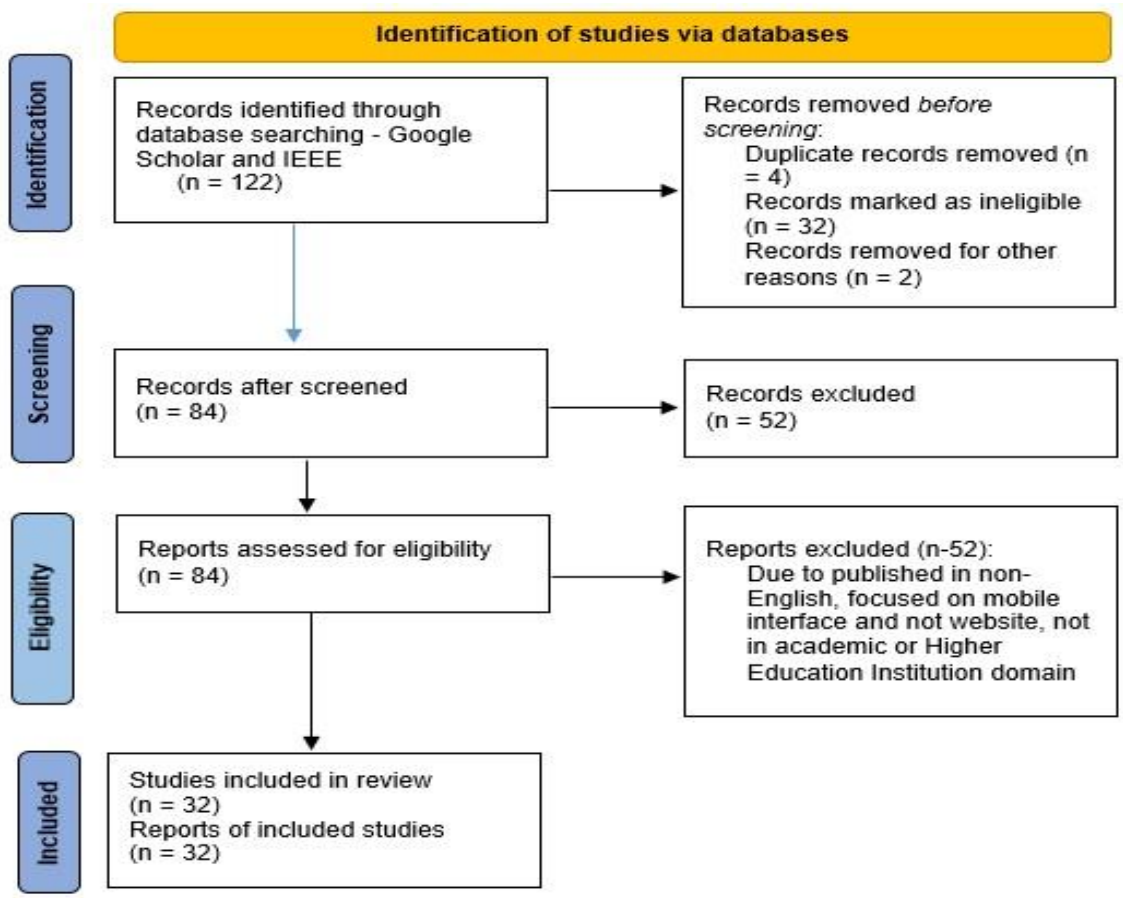
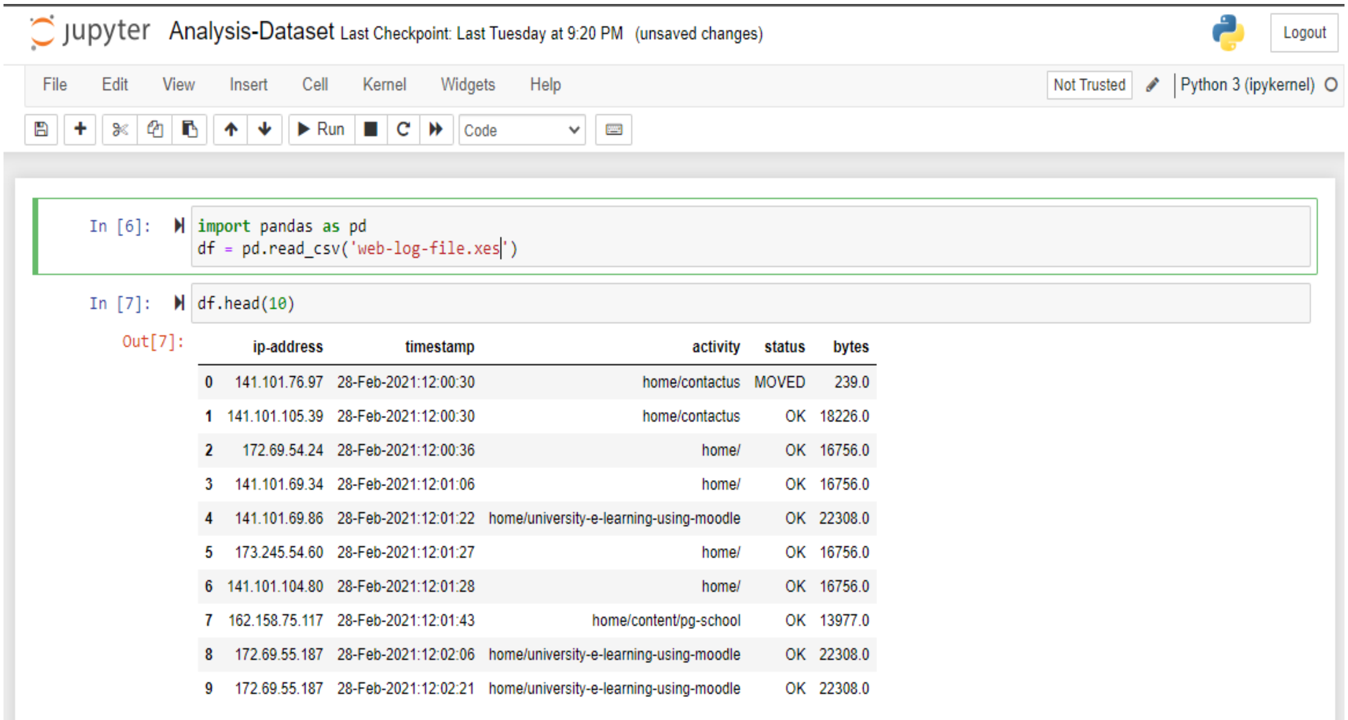


Fig. 5 PRISMA stages adopted in the study

A. Results of Data Cleaning, Preprocessing, and Transformation

Figure 6 shows the screenshot of the importation of the event log data using the panda package by calling the command *pd.read_csv ('web-log-file. xes')* as a variable named df into the python jupyter notebook resident in the Anaconda environment. Following the data importation using the panda, the first 10 records in the event log data were displayed using the command *df. Head (10)* shows the contents of the first ten records stored in the dataset.



	ip-address	timestamp	activity	status	bytes
0	141.101.76.97	28-Feb-2021:12:00:30	home/contactus	MOVED	239.0
1	141.101.105.39	28-Feb-2021:12:00:30	home/contactus	OK	18226.0
2	172.69.54.24	28-Feb-2021:12:00:36	home/	OK	16756.0
3	141.101.69.34	28-Feb-2021:12:01:06	home/	OK	16756.0
4	141.101.69.86	28-Feb-2021:12:01:22	home/university-e-learning-using-moodle	OK	22308.0
5	173.245.54.60	28-Feb-2021:12:01:27	home/	OK	16756.0
6	141.101.104.80	28-Feb-2021:12:01:28	home/	OK	16756.0
7	162.158.75.117	28-Feb-2021:12:01:43	home/content/pg-school	OK	13977.0
8	172.69.55.187	28-Feb-2021:12:02:06	home/university-e-learning-using-moodle	OK	22308.0
9	172.69.55.187	28-Feb-2021:12:02:21	home/university-e-learning-using-moodle	OK	22308.0

Fig. 6 Screenshot of the importation of the event log data

After removing records with missing data from the event log data, the status codes that had been converted into their respective HTTP server responses were invested. This was done using the command *df['status'].value_counts()* which displays the number of records that contain the respective HTTP server responses to their respective status codes. Out of the 418194 records in the event data log, 304052 were discovered to have a server response value of OK. As a result, the records without server responses of OK were removed from the dataset, leaving behind a total of 304052 records, which accounted for 34.57% of the original dataset. Figure 7 shows a screenshot of the frequency of the HTTP server requests as found in the event data log. Table 2 summarizes the number of records in the event data log following each preprocessing stage performed on the dataset.

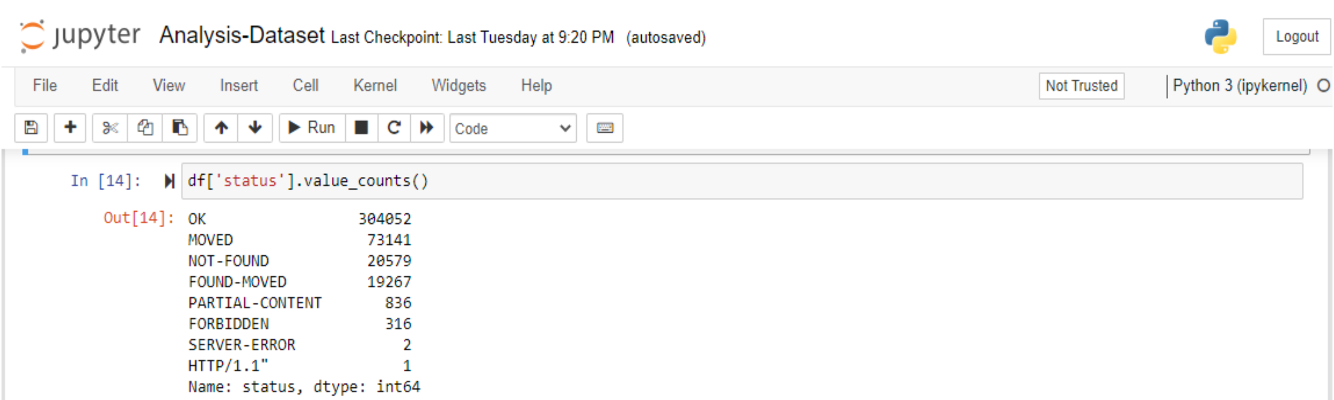


Fig. 7 Screenshot of the distribution of HTTP server response

Table 2. Results of data cleaning and preprocessing

Activity	Number of Records	Proportion (%)
Original Data	879475	100.00
After data cleaning	418194	47.55
OK, status records left	304052	34.57

B. Result of Analysis of Event Data

During the analysis, five process mining algorithms were applied to the event log data adopted in this study. However, as a pilot test of the feasibility of the approach considered, a portion of the dataset was used to assess the performance of the approach considered

In this study. Following the application of the various process mining algorithms considered in this study, the performance of the algorithms was evaluated based on four (4) metrics. Table 3 shows the performance evaluation results of the process mining algorithms considered in this study.

Table 3. Results of the performance evaluation of process mining algorithms

Algorithm	Simplicity	Fitness	Precision	Generalization
Alpha Miner	-	0.674	0.727	0.986
Heuristic Miner	-	1.000	0.000	0.000
ILP Miner	-	0.111	0.385	0.922
Inductive Miner	1.000	0.903	0.231	0.966
Evolutionary Tree Miner	0.611	0.994	0.303	0.977

The evaluation of the performance of the five (5) process mining algorithms revealed remarkable results. The results revealed that among the algorithms considered in this study, the inductive and evolutionary miner were the only algorithms evaluated based on all four (4) performance metrics. There were no recorded values for assessing the remaining three (3) algorithms based on simplicity.

The simplicity of the evolutionary miner revealed that it could focus on the important aspects of the underlying process by 61.1%; however, the inductive miner was able to perfectly focus on the important aspects of the underlying process with a proportion of 100%. Incidentally, the heuristic miner was the only algorithm that perfectly represented the underlying behavior of the event data log with the value of 100% but failed to achieve anything based on precision and generalization. As a result, the heuristic miner was unable to express that behaviour that was not evident in the event log nor show any potential of predicting the future behaviour of the underlying process.

However, the inductive and evolutionary miners represented the underlying behaviour present in the event log data by a proportion of 90.3% and 99.4%, respectively. Unlike the heuristic miner, the generalization of the other four (4) models revealed that they showed the potential of predicting the future behaviour of the process owing to their generalization of over 90%. However, the alpha miner proved to have the greatest potential with 98.6%, followed by the evolutionary miner with a value of 97.7%. The heuristic miner was also able to demonstrate the best capacity at identifying the behavior which is not seen in the event log among the algorithms with a proportion of 72.7%, with the inductive miner doing very poorly with a proportion of 23.1%. Figure 8 shows a bar chart distribution of the performance of the algorithms used for generating the process model of the event data log considered in this study.

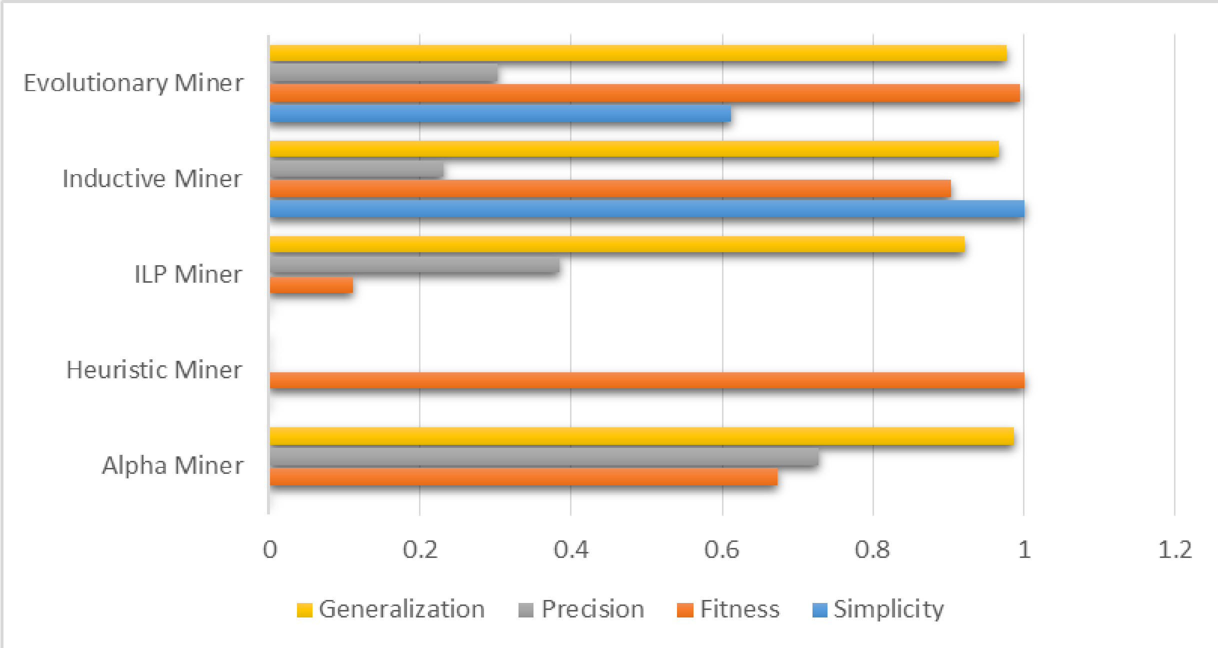


Fig. 8 Bar chart distribution of performance evaluation

Appendix A shows a screenshot of the Petri net model for the underlying process extracted from the event data log using the inductive miner algorithm. This was done by the inductive miner using a divide-and-conquer approach. The process involved repeatedly splitting the event log into sub-logs until it finds the most likely split; the split method gives a structured process model. As shown in the figure, the white squares represent the transitions representing an action or activity; actions represent the links/pages visited. The black squares are silent transitions with no observable activity; they act as placeholders to connect active transitions. Circles represent a place where the tokens are placed; the edges between nodes are arcs represented by arrows. The model shows the control flow of the activities. Appendix B extracts the control flow model extracted from the process model indicated by the red square embedded in Figure 9. As shown in the diagram, each path along the process model is a potential link that can be integrated into the website as a potential path of interest to a user visiting the website. By adopting this mechanism, users are more likely to spend less time looking for information, thereby being motivated to spend more time exploring the website.

V. DISCUSSIONS

Among the four (4) process mining algorithms that were adopted in this study for the extraction of the underlying process, which describes the navigational behaviour of website users based on information stored on the event data log, it was observed that the inductive miner proved to be the most effective model. This is because the process model developed using the inductive miner proved to identify the most important aspects of the process from the event data. This was also evident in its ability to represent the underlying behaviour of the event data log, thus providing insight and clarity, in addition, to its ability to prove useful in being used for future predictive behaviour of the underlying process.

VI. Conclusion

This study revealed that the end-to-end process of website users could be discovered based on information extracted from web server log files. By adopting the process mining modeling techniques generated from website activities by users, it is possible to extract hidden insights that are more detailed and objective than what existing website usability methods will show. Using a process-driven approach for website usability evaluation extends the capability of existing methods as it shows more detailed information about how the users are using the website. Future studies will focus on extracting the relationships extracted from the patterns generated from the user navigational behaviour. These relationships will form a basis for integrating rule-based recommender systems based on the patterns extracted from the process mining model.

REFERENCES

[1] International Organization for Standardization., Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts (ISO 9241-11:2018). Retrieved from https://www.iso.org/standard/63500.html. (2018).

[2] Hartson, R., & Pyla, P., Background: Introduction. The UX Book, (2019)95–113. https://doi.org/10.1016/b978-0-12-805342-3.00006-0

[3] Lee, H., Ka-hyun Lee, K., & Choi, J., A Structural Model for Unity of Experience: Connecting User Experience, Customer Experience, and Brand Experience. Journal of Usability Studies, 14(1) (2018) 8–34. Retrieved from https://uxpajournal.org/wp-content/uploads/sites/8/pdf/JUS_Lee_Nov2018.pdf

[4] Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics, 76(November 2018) (2020) 102932. https://doi.org/10.1016/j.ergon.2020.102932

[5] Almeraj, Z., Boujarwah, F., Alhuwail, D., & Qadri, R., Evaluating the Accessibility Of Higher Education Institution Websites in the State of Kuwait: Empirical Evidence. Universal Access in the Information Society, 20(1) (2021) 121–138. Https://Doi.Org/10.1007/S10209-020-00717-8

[6] Ramanayaka, K. H., Chen, X., & Shi, B., UNSCALE : A Fuzzy-based Multi-criteria Usability Evaluation Framework for Measuring and Evaluating Library Websites UNSCALE : A Fuzzy-based Multi-criteria Usability Evaluation Framework for, 4602 (2018). https://doi.org/10.1080/02564602.2018.1498032

[7] Atta-ur-Rahman, Dash, S., Luhach, A. K., Chilamkurti, N., Baek, S., & Nam, Y., A Neuro-fuzzy approach for user behaviour classification and prediction. Journal of Cloud Computing, 8(1) (2019). https://doi.org/10.1186/s13677-019-0144-9

[8] Keshavarz, H., & Givi, M. E., Website Evaluation Frameworks: IS oriented vs. Business Oriented Models. 2020 6th International Conference on Web Research, ICWR, (2020) 223–228. https://doi.org/10.1109/ICWR49608.2020.9122316

[9] Augusto, A., Conforti, R., Armas-Cervantes, A., Dumas, M., & La Rosa, M., Measuring Fitness and Precision of Automatically Discovered Process Models: A Principled and Scalable Approach. IEEE Transactions on Knowledge and Data Engineering, 4347(c) (2020) 1–1. https://doi.org/10.1109/tkde.2020.3003258

[10] Eryk, L., Introduction to Process Mining. Learn the basics of process mining and… | by Eryk Lewinson | Towards Data Science. Retrieved November 28, (2020). From https://towardsdatascience.com/introduction-to-process-mining-5f4ce985b7e5

[11] Ghasemi, M., & Amyot, D., A systematic literature review of goal-oriented process mining from event logs to goals. Requirements Engineering, 25(1) (2020) 67–93. https://doi.org/10.1007/s00766-018-00308-3

[12] Graafmans, T., Turetken, O., Poppelaars, H., & Fahland, D., Process Mining for Six Sigma. Business & Information Systems Engineering. (2020). https://doi.org/10.1007/s12599-020-00649-w

[13] Shevchenko, L., Analysis of library website users' behavior to Optimize Virtual Information and Library Services. Journal of Information Science Theory and Practice, 8(1) (2020) 45–55. https://doi.org/10.1633/JISTaP.2020.8.1.4

[14] Anyaoku, E. N., & Akpojotor, L. O., Usability Evaluation of University Library Websites in South-South Nigeria. Library Philosophy and Practice, 2020(January) (2020) 1–26.

[15] Manzoor, M., Hussain, W., Sohaib, O., Hussain, F. K., & Alkhalaf, S., Methodological Investigation for Enhancing the Usability of University Websites. Journal of Ambient Intelligence And Humanized Computing, 10(2) (2019) 531–549. Https://Doi.Org/10.1007/S12652-018-0686-6

[16] Haritha, S., Vismitha, P. S. R. I., Kavyanjali, S., & Shireen, S. ., Analysis of User's Behavior I Structured E-Commerce Websites, 1(9) (2018) 206–211.

[17] Adedeji, F. ., Adekunle, Y. ., Adebayo, A. ., Alao, O. ., & Akande, O., A Systematic Review on Usability Evaluation for University Websites. International Journal of Computer Applications Technology and Research, 11(2) (2022) 22–28. https://doi.org/10.7753/ijcatr1102.1003

[18] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Grp, P. ., Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). Physical Therapy, 89(9) (2009) 873–880. https://doi.org/10.1371/journal.pmed.1000097

[19] Nwasra, N., Basir, N., & Marhusin, M. F, Evaluation of Malaysian Universities Websites based on Quality in Use Evaluation Model, 8(4) (2018) 1417–1422.

[20] Park, J., Han, S. H., Kim, H. K., Cho, Y., & Park, W., Developing Elements of User Experience For Mobile Phones and Services: Survey, Interview, and Observation Approaches. Human Factors and Ergonomics In Manufacturing, 23(4) (2013) 279–293. Https://Doi.Org/10.1002/Hfm.20316

[21] Buijs, J. C. A., Flexible Evolutionary Algorithms for Mining Structured Process Models. 2014. Eindhoven. (2014) Https://Doi.Org/10.6100/IR780920

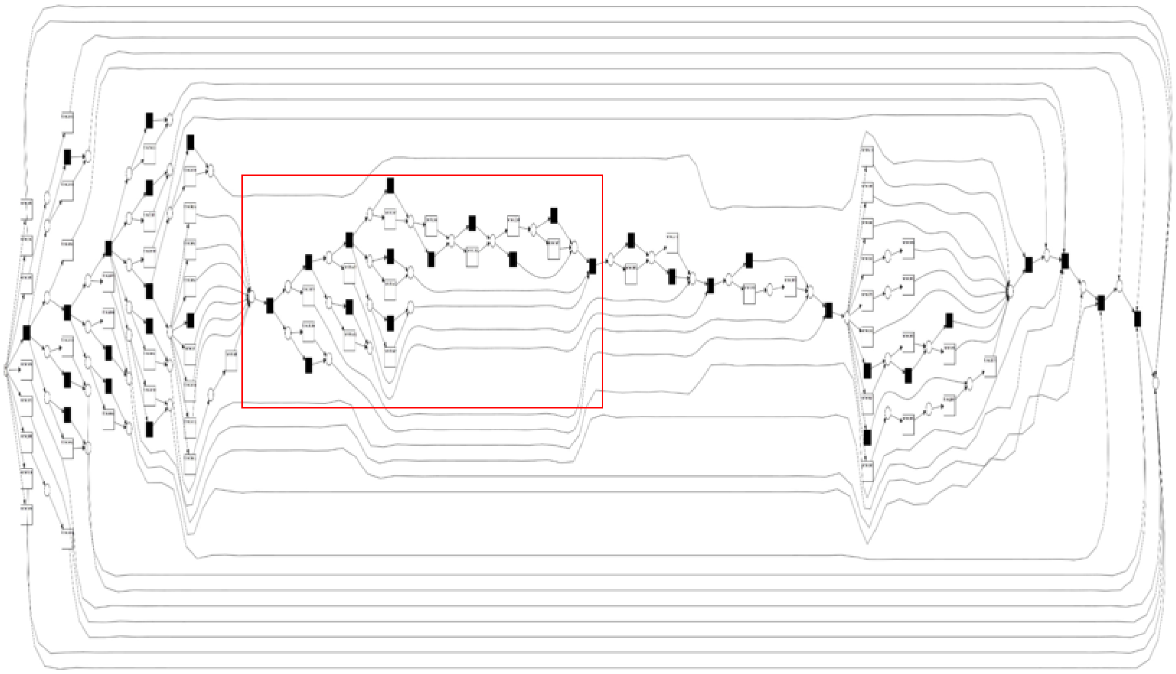
[22] Grigorova, K., Malysheva, E., & Bobrovskiy, S., Application of Data Mining and Process Mining Approaches for Improving E-Learning Processes. CEUR Workshop Proceedings, 1903 (2017) 115–121. https://doi.org/10.18287/1613-0073-2017-1903-115-121

[23] Bogarín, A., Cerezo, R., & Romero, C., A Survey on Educational Process Mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(1) (2018a). https://doi.org/10.1002/widm.1230

[24] Van der Aalst, W., Process mining: Data science in action. Process Mining: Data Science in Action. (2016). https://doi.org/10.1007/978-3-662-49851-4

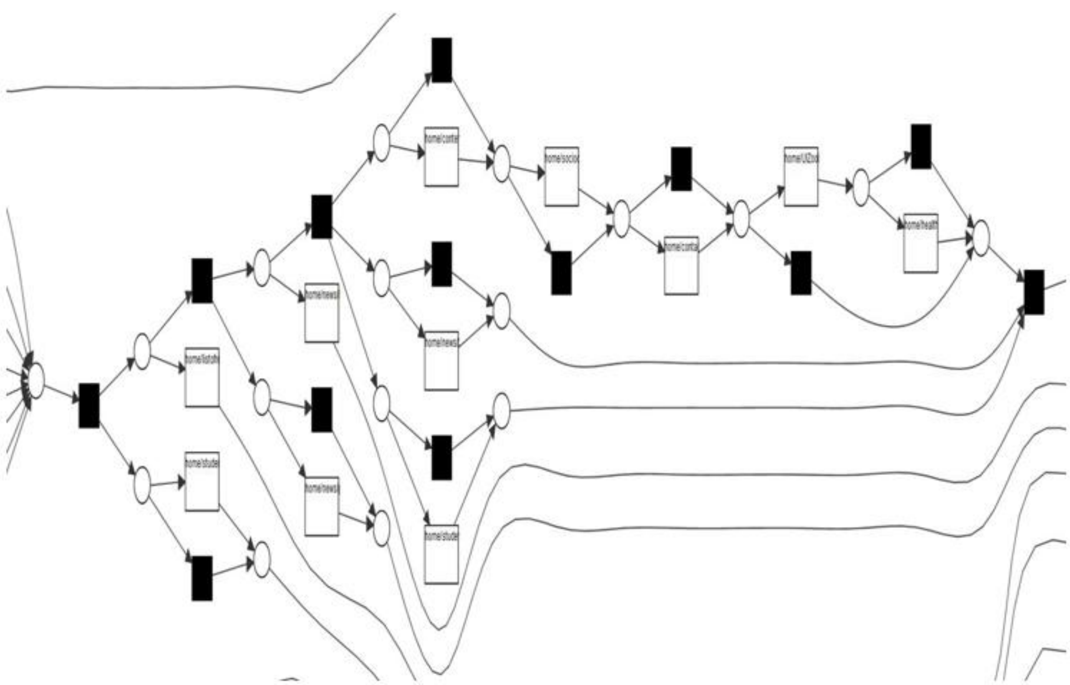
[25] Deedam, F. B., Thomas, E., & Taylor, O. E., Accessibility and Usability Evaluation of State-Owned Universities Website in Nigeria. International Journal of Engineering Trends and Technology, 56(1) (2018) 31–36. https://doi.org/10.14445/22315381/ijett-v56p206

APPENDIX A



A screenshot of the process model generated by the inductive miner from the event data log

APPENDIX B



An extract from the process model generated by the inductive miner algorithm.