International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
Article Received: *25 October 2020* ***Revised:*** *12 November 2020* ***Accepted:*** *30 December 2020*

Artificial Intelligence-Based Approaches for Enhanced Web Personalization: Transforming User Experience and Adaptive Web Interactions

Laxmi Choudhary[1], Shashank Swami[2*]

[1]Research Scholar, Department of Computer Science, Sabarmati University, laxmi.choudhary23@gmail.com
[2*]Professor, Department of Computer Science, Sabarmati University, Swamishashank78@gmail.com

Abstract

Web usage mining is essential for getting user behavior out of Weblogs and customizing business Websites. This work aims to discover these hidden weblog rules by implementing the Apriori Prefix Tree (PT) algorithm on the PSNBC data set. The most important task is to predict the next page based on the user's web activity. The performance of the generated rules was measured with the help of fundamental factors including lift, confidence and support. The findings show that confidence does not vary across the pages but that lift and support correlate highly with the page's significance. Most visited web pages such as News, Front Page, On-air News, Sports and BBS received higher traffic than commonly visited pages like Travel, PSN-News and PSN-Sports. From these findings, it becomes evident that there is potential in web usage mining in consequent log analysis from servers in yielding insightful knowledge for analysis of user conduct and generating personalization content.

Keywords: Web Usage Mining, Apriori Prefix Tree Algorithm, Web Personalization, User Behavior Prediction, Lift Metric, Confidence Metric, Support Metric, PSNBC Dataset, Web Page Engagement

Introduction

Specifically in the context of the last years, research interest has risen substantially in understand users activity online (Chakrabarti, 2002). This trend arises from the recognition that increasing a website's value for users requires more than just adding content; it also involves delivering information in a timely, accessible and user-friendly manner. Studies indicate that the time spent online is increasing steadily as user activities on websites become more streamlined and interactive '(Hung et al. (2013); Phoa & Sanchez (2013); Anitha & Krishnan (2011))'. Current estimates reveal over 3 billion internet users worldwide, with Asia alone accounting for 1.38 billion users. While many users are non-experts in digital technologies, they recognize the significant role of the internet in their daily lives. With web activity constantly growing, online platforms are under increasing pressure to manage and organize content effectively. Research shows that over 1,200 GB of web modifications are made monthly and millions of new pages are added to the internet daily. Remarkably, new web servers go live every two hours, resulting in an estimated one webpage for every two individuals globally.

Given this vast and dynamic environment, understanding user behavior has become essential for personalizing web experiences. The goal of this study is to enhance visitor prediction models by applying data mining techniques to forecast which pages will attract user attention '(Chandra & Basker, 2000; Sanchez & Liu, 2011; Liu & Peng, 2013)'. Web usage mining is concerned with the discovery of more patterns from massive databases to support the requirements of the web user (Robert et al., 1999; Suraya et al., 2011; Jacob et al., 2013). Most of these techniques integrate the findings of artificial intelligence, machine learning, statistics and database management (Kotsiantis et al., 2007; Han & Kamber, 2011; Ramani & Jacob, 2013a,'b). Web usage mining identifies the visitors and their activities within websites and other related systems from various sources (Chunsheng & Li, 2014). These include:

- Web Server Logs: HTML rules containing the recording of users' interactions with websites.
- Commercial Application Servers: Other online business applications like Web Logic Story Server that retain the user activity information '(Madhuri, 2002; Yang et al., 2009; Wen-Hai, 2010; Veeramalai et al., 2010)'.

Web usage mining is most useful for businesses and organizations since it offers directions on what their users are doing and how to manage web applications well. One traditional way to look for relationships between variables in large databases is Association Rule Mining (ARM) (Agrawal & Srikant, 1994; Kotsiantis &

Kanellopoulos, 2001; Kumar & Chezian, 2012). ARM is extensively implemented to discover patterns between variables in datasets using metrics such as support, confidence and lift. The presented approach is based on PSNBC dataset logs drawn from users activity and relies on the Apriori algorithm, the association rule mining algorithm. The Apriori algorithm used in the current study expands the concepts regarding frequent items. It makes use of the diagram, which ensures that the support is closed downward effectively for rule generation, which is an efficient way, as supported by 'Wang et al., 2000; Renáta & Vajk, 2006'. In the subsequent sections, we explain how, through the Apriori PT procedure we proposed, it is possible to penetrate which page a given user will probably be browsing next by the frequency of visited page utilization and interest.

Literature Reviews

This section presents Web usage mining studies to organize the increasing number of publications in the field and recent developments.

Zhang and Chen (2014) used association rule mining to present a forensic model of web browsing behavior. The study then converted log files into transactional database, decomposing the web log data into single web pages visited and using an Apriori algorithm to generate the frequent web usage patterns. This aided in identifying anonymous and unknown behaviors on the World Wide Web.

Tassa (2014) suggested a way to mine association rules in databases that are horizontally partitioned while maintaining privacy. This system used multi-party algorithms to increase data security when mining rules on disparate datasets. As for computational and communication efficiency, Tassa's protocol was less complex than the previous methods.

Mary and Malarvizhi used the PSNBC dataset to recommend a weighted Apriori-based system in 2012. The study provided better recommendations for web pages with 35% support and 64% confidence from web usage traces compared to traditional rule-mining models. I also used Bayesian hierarchical models similar to Sanchez and Liu (2011) to group user sessions and model the page visitation.

The PSNBC dataset, for the first time employed by Suraja et al. (2011), contains 17 URL categories and nearly 989,818 users while proposing the personalized minimum support (P_minsup) criterion. Their system was another system that adopted the SPADE algorithm to keep discovering common patterns to enhance the prediction of the page sequences.

Suresh et al. (2011) used an improved fuzzy c-means clustering technique to extract weblog navigational behavior patterns. Their research proved enhanced clustering performance when processing PSNBC web data and a better understanding of users browsing habits.

Santhisree & Damodaram (2010) proposed OPTICS for clustering web usage data based on density. The OPTICS algorithm was used to determine the intra and inter-cluster distances and ship them to create 5-dimensional data describing a user's behavior and where geometric and other types of distances were used, including Euclidean and Jaccard distances. The findings offered functional patterns for forecasting future visits. This work extends these methodologies by extracting association rules with the Apriori method and generate the user behavior of the PSNBC data set. This framework and the findings are presented in the subsequent sections of this paper.

Methodology

These include the general research methodology, the PSNBC data collection methods and data pre-processing and analysis methods employed in inferring user behavior patterns.

1. Research Design

In this study, association rule mining— a type of online usage mining— is combined with a descriptive research design. Specifically, the Apriori algorithm is applied to weblogs to estimate the behavior of users according to the frequently obtained itemsets.

2. Data Collection

- **Primary Data:**

Data is sourced from Internet Information Server (IIS) logs of psnbc.com and psn.com for a single day, September 28, 1999.

These logs capture 989,818 user records across 17 categorized web pages, including front page, sports, news and tech (Chandra & Basker, 2000).

- **Secondary Data:**

Literature and algorithmic references such as Agrawal & Srikant (1994) and other studies on web usage mining.

3. Data Pre-Processing

- **Format:** Excel was used to organize the data, with each row denoting a distinct user and each column denoting a category of web page.
- **Data Cleaning:** TANAGRA (an open-source data mining tool) was used to import and clean the data for accuracy. Missing data or corrupted entries were excluded from the analysis.
- **Transformations:** The raw data (user logs) was converted into transactional data for association rule mining. Each user's sequence of page visits was categorized into itemsets to prepare for pattern extraction.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
***Article Received:** 25 October 2020 **Revised:** 12 November 2020 **Accepted:** 30 December 2020*

4. Association Rule Mining using the Apriori Algorithm

- **Algorithm Overview:** Using support and confidence thresholds, the Apriori algorithm creates association rules and finds frequently occurring itemsets (Rahman et.al 2019).
- **Stages of the Apriori Algorithm:**

a) Join Operation: To create candidate k-itemsets, frequent (k-1)-itemsets are linked.
b) Prune Operation: Infrequent itemsets are eliminated based on the minimum support threshold **(**Prithiviraj & Porkodi 2015).

• Pseudocode Implementation:

Ck: Candidate itemset of size k
Lk: Frequent itemset of size k
L1 = {frequent items};
for (k = 1; Lk ≠ ∅; k++) do begin
Ck+1 = candidates generated from Lk;
foreach transaction t in the database do increment the count of all candidates in Ck+1 contained in t;
Lk+1 = candidates in Ck+1withmin_support;
End
return Uk Lk;

5. Thresholds Used for Rule Generation

- Support Threshold: 0.6%
- Confidence Threshold: 100%

The Apriori-prefix tree (PT) variant was applied for efficient handling of large datasets, ensuring quick rule generation.

6. Data Analysis Techniques

- **Association Rule Mining:**
  Rules were generated based on frequent itemsets with the specified support and confidence levels. Patterns were examined to predict the next likely page a user will visit based on browsing behavior (Patil & Gupta 2017).
- **Descriptive Statistics:**
  Frequency distributions were calculated to identify commonly visited page categories. The analysis also involved evaluating rule lift values to determine the significance of patterns.

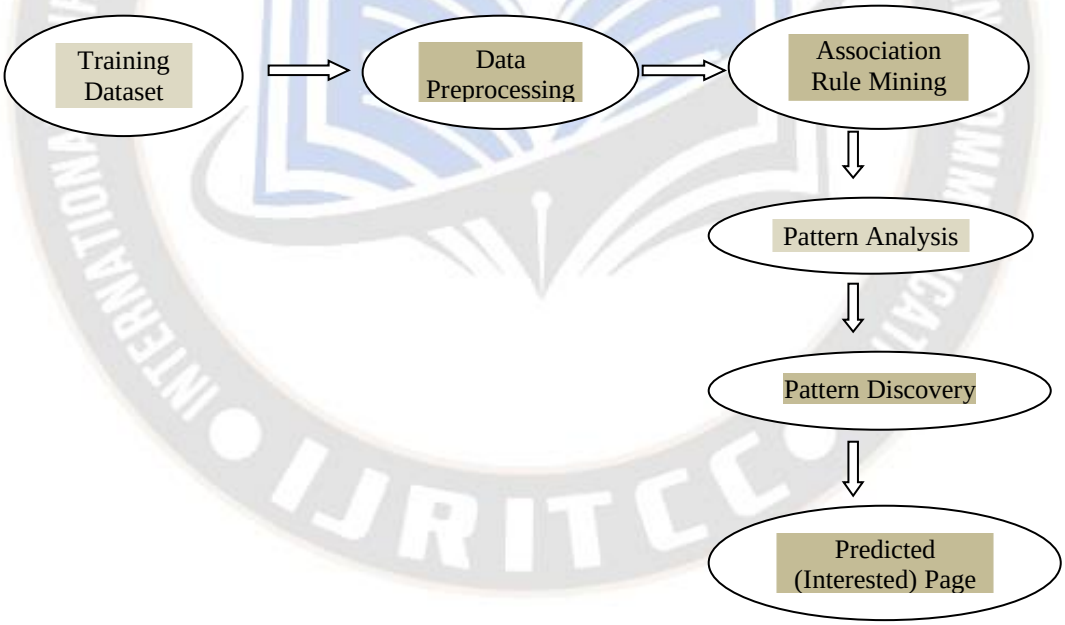

Figure 1: Web Navigation Prediction Framework for Web Page Recommendation

This section reports the results of mining the PSNBC dataset using association rule mining. It contains frequency analysis, selected rule's metrics identification and how often a user is predicted to return. Moreover, tabular and graphical forms are applied to present the results.

7. Pattern Analysis of User Behavior

The number of pages in each cycle and the sequence of their visits were used to generate rules. For instance, a 3-sequence itemset might appear as P1 | P2 | P3, while a 4-sequence itemset could be P2 | P4 | P5 | P8. Itemsets with fewer than three pages were excluded due to their limited relevance, as they generally indicate less engaged browsing behavior (Anitha & Krishnan 2011).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
Article Received: *25 October 2020* ***Revised:*** *12 November 2020* ***Accepted:*** *30 December 2020*

Table 1: Regulations and Ordering Tables Relying on Antecedents

S. No	Sequence Itemset	Number of Rules	Min. Support
A	3	48	0.00001720
B	4	721	0.00006850
C	5	6304	0.00002910
D	6	26181	0.00003630
E	7	54126	0.00020310
F	8	88155	0.00007860
G	9	102034	0.00002910
H	10	82086	0.00008650
I	11	39689	0.00006080
J	12	22183	0.00006850
K	13	7020	0.00003330
L	14	1684	0.00001380
M	15	118	0.00001210
N	16	6	0.00001210
O	17	10	0.00001210

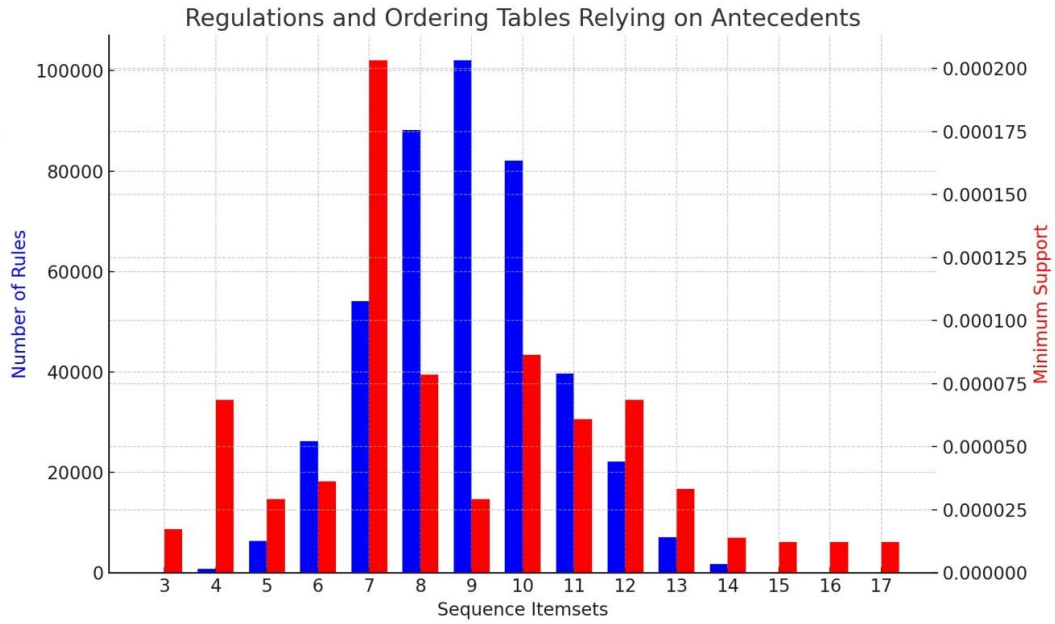

Figure 2: Number of Rules Generated for Each Sequence Itemset Along with Minimum Support

Key Insights from the Results

a) Number of Rules Grows with Sequence Length:
As the sequence itemsets become longer (moving from 3-itemsets to 9-itemsets and beyond), the number of generated rules increases significantly. This indicates that longer browsing patterns provide more meaningful insights into user behavior. 9-sequence itemsets generated the highest number of rules (103,134), showing that users often follow longer, multi-step navigation paths when browsing.

b) Minimum Support Thresholds:

The minimum support threshold is a parameter that establishes the minimum number of times an itemset needs to appear in the dataset in order to be deemed important. Smaller itemsets (e.g., 3-sequence) have lower minimum support thresholds since their frequency is limited. Larger itemsets (e.g., 7or 9-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
Article Received: *25 October 2020* ***Revised:*** *12 November 2020* ***Accepted:*** *30 December 2020*

sequence) exhibit higher thresholds, as these sequences capture more consistent user behavior.

c) Graph Interpretation – Number of Rules and Minimum Support

- The graph's blue bars show how many rules are produced for every sequence in the itemset.
- Red dashed lines show the minimum support values corresponding to those itemsets.

The graph shows an evident relationship between sequence length and rules generated. With an increase in the sequence length, there is a proportional increase in the minimum support to warrant capturing all those that exist.

Experimental Results

In this section provides the results across three major components:

1) Association rule metrics highlight the key measures used for evaluating the rules.
2) Rule antecedents are when student actions under consideration are characterized by specific patterns such as the sequence of actions a user performs.
3) Rule consequents examine the consequent outcomes & anticipated future use of the QAM by the user.

1. Association Rule Metrics

Association rule mining evaluates patterns based on three primary metrics: Support, Confidence and Lift. These metrics are used to measure the degree and importance of the rules.

a) Support: The support (X) of itemset is the % of transactions in the dataset that contain an itemset.

$$\text{Support} = \frac{\text{Number of transactions containing X}}{\text{Total transactions}}$$

b) Confidence: Confidence approximates the probability of occurrence of itemset Y contingent upon the occurrence of set X. It depicts the level of link between X and Y.

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

c) Lift: Lift measures the level of dependency between itemsets X and Y concerning each itemset's support. This implies that as Y increases, X will likewise grow in value and vice versa. The value of 'Lift' is more significant than 1, which means the association between X and Y is positive.

$$\text{Lift}(X \Rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X) \times \text{Support}(Y)}$$

2. Rule Consequent Analysis

This table presents the number of rules generated concerning each page, and lift values control the significance of the specific pages regarding user engagement.

Table 2: Rules and Itemset Sequences Determining Rule Consequences

S.no	Page	Number of Rules	Lift
A	P1	3807	0.3123
B	P6	3807	0.2071
C	P2	3807	0.1550
D	P3	3072	0.1130
E	P4	3807	0.1120
F	P14	3807	0.1090
G	P12	3807	0.1030
H	P8	3807	0.0830
I	P9	3072	0.8110
J	P7	3072	0.0710
K	P13	3072	0.0670
L	P11	3807	0.0530
M	P10	3807	0.0410
N	P5	3807	0.0230

i. Findings:

The pages that received the most lift was P1 and P6 which are considered landing pages or very important for the user's movements.

The feature Page P9 was exciting as it showed a high lift of 0.811, indicating a good predictability of user page visits.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
Article Received: *25 October 2020* ***Revised:*** *12 November 2020* ***Accepted:*** *30 December 2020*

3. Rule Antecedent Analysis

The following section elucidates the rule antecedents. Preceding examples of three or four dwell pages are used to analyze the subsequent behavioral output and discover end-user patterns (Debahuti 2010).

Table 3: Pattern Discovery from Rule Antecedents in Three-Sequence Itemsets

Rule No.	Rule	Highest Lift	Predicted Subsequent Page
R1	P16 | P17 | P12	0.3123010	P1
R2	P16 | P17 | P14	0.2071121	P6
R3	P16 | P17 | P6	0.3123010	P1
R4	P16 | P17 | P7	0.3123010	P1

i. Findings:

Based on several sequences, R1, R3, R4, P1 remains popular which gives an indication that most users revisit this destination.

P6 remains another frequently viewed page, it is accessed after particular call sequences R2 and therefore, is important addressing various paths.

4. Summary of Results

i. Rule Generation:

The study described sequences of pages within which rules were produced. Itemsets containing less than three pages were omitted as the occurrence of such patterns were relatively insignificant.

ii. Impact of Page Importance:

Pages with higher lift values (such as P1 and P9) play a critical role in predicting future actions, demonstrating that users are more likely to visit these pages.

iii. Frequent Pathways:

Sequences like P16 | P17 | P12 → P1 reveal that users tend to navigate through specific categories and returning to primary pages. These findings demonstrate the effectiveness of using association rule mining to predict user behavior and optimize the browsing experience on websites for web data.

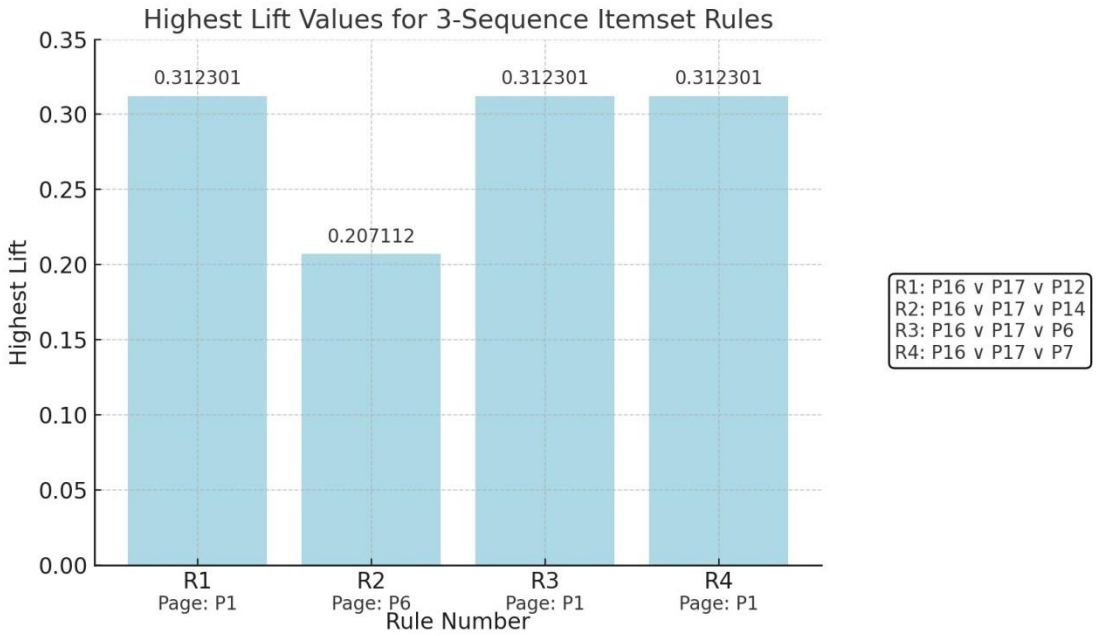

Figure 3: Web Page Prediction From 3-Sequence Itemsets

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
Article Received: *25 October 2020* ***Revised:*** *12 November 2020* ***Accepted:*** *30 December 2020*

Table 4: Pattern Discovery From Rule Antecedents in Four-Sequence Itemsets

Rule No.	Rule	Highest Lift	Predicted Subsequent Page
R1	P16|P17|P13|P9	0.3123010	P1
R2	P16|P17|P13|P10	0.3123010	P1
R3	P16|P17|P13|P12	0.3123010	P1
R4	P16|P17|P13|P4	0.2071121	P6
R5	P16|P17|P13|P3	0.2071121	P6
R6	P16|P17|P7|P9	0.3123010	P1
R7	P16|P17|P6|P5	0.3123010	P1
R8	P16|P17|P6|P3	0.3123010	P1
R9	P16|P17|P7|P4	0.3123010	P1

iv. Rule Antecedent Analysis

The analysis of rule antecedents began with three-sequence itemsets and extended up to six-sequence itemsets. As the number of sequences itemsets increased, the frequency of rules generated steadily declined, suggesting that higher sequence patterns attracted fewer visitors. Based on lift values, Table 3 displays the most significant three-sequence itemsets. Single-occurrence limitations were not stressed for three-sequence itemsets due to the low number of visits. Rule R1 (P16 **|** P17 | P12) represents the itemset having the lowest support (0.0000172), occurring 9 times on different pages, while the item with the maximum lift (0.3123010) focused on P1 as the projected upcoming page. In a similar vein, rule R2 (P16 **|** P17 | P14) was seen six times and P6 was predicted to be the following page with the biggest lift (0.2071121).

Additionally, rule R3 (P16 **|** P17 | P6) appeared six times, showing a lift value of (0.3123010), also predicting P1 as the next page. Rule R4 (P16 **|** P17 **|** P7) shared the same lift value (0.3123010) and similarly predicted P1 as the subsequent access page. These findings suggest that P1 and P6 emerge as frequent next destinations following sequences involving pages like PSN-News (P16) and PSN-Sports (P17) (Hung et.al 2013).

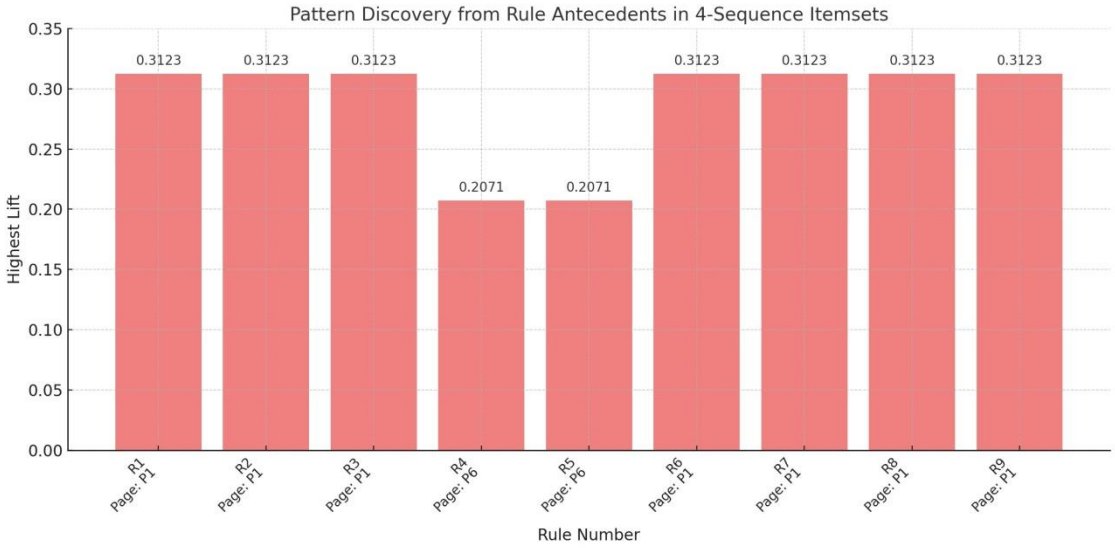

Figure 4: Web Page Prediction From Four-Sequence Itemsets

The analysis of the rule antecedents reveals key patterns across both three-sequence and four-sequence itemsets. Below are the primary findings:

a) Decline in Rule Generation with Increased Sequence Length:

As the number of pages in the sequence increases from three to six, the frequency of rules decreases. This suggests that longer browsing patterns are less common,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
***Article Received:** 25 October 2020 **Revised:** 12 November 2020 **Accepted:** 30 December 2020*

possibly due to reduced user engagement over extended sessions.

b) Prominence of Key Pages (P1 and P6):

Pages like P1 (Front Page) and P6 (On-Air) frequently appear as the predicted subsequent destinations, indicating these pages play a significant role in user navigation across multiple browsing sessions. For example (Kum et.al 2005):

- **Three-Sequence Itemsets:**
  Rules involving P16 (PSN-News) and P17 (PSN-Sports) often lead to P1 as the next visited page.
- **Four-Sequence Itemsets:**
  P1 continues to dominate as the predicted subsequent page, especially when sequences involve P15 (Travel)**,** P8 (Summary) or other pages.

c) High Lift Values Indicating Strong Predictive Power:

The highest lift value (0.3123010) is consistently associated with sequences predicting P1, suggesting a strong association between browsing patterns and P1 as a key access point. This reflects user preference to visit or return to the Front Page frequently during their browsing sessions.

d) Multiple Pages Funnel Users to P1 and P6:

Across both the three-sequence and four-sequence itemsets, rules such as P16 **|** P17 **|** P6 and P16 **|** P17 ∧ P13**|** P9 frequently predict P1. This indicates that multiple browsing pathways converge to these prominent pages, signifying their central role in the website's navigation structure.

e) Role of News and Sports Pages in Predicting Navigation:

Many rules involve PSN-News (P16) and PSN-Sports (P17) as initial or intermediary pages, highlighting their importance in driving traffic to other significant pages like the Front Page (P1) or On-Air (P6). These pages serve as entry points or transitional pages within the broader user journey.

These findings emphasize that user navigation on the website revolves around a few key pages. P1 (Front Page) emerges as the most frequently visited destination, reinforcing its importance in the user experience. Similarly, P6 (On-Air) plays a significant role, especially in longer browsing sequences, indicating it as a secondary hub in the website's architecture.

f) Analysis of Multi-Sequence Itemsets and Predicted Page Navigation Patterns:

The analysis focused on identifying the next probable page based on the highest lift values obtained from various sequence itemsets. For three-sequence itemsets, the goal was to find the subsequent page with the highest lift, where a lift value of 0.3123010 was observed in multiple instances. The 'lift' metric was instrumental in predicting the next page in each case (Gao et.al 2009). In several patterns (R1, R3 and R4), users transitioning from PSN-News (P16) to PSN-Sports (P17), followed by Summary (P13)**,** Wealth (P8) or Health **(**P9**)**, were likely to visit the Front Page (P1) next. In other cases, such as Rule R2**,** the data suggested that after visiting PSN-Sports (P17) and PSN-News (P16)**,** users were likely to reach the On-air (P6) page. Across multiple three-sequence itemsets, pages P1 and P6 consistently emerged as the most probable next pages compared to other options. Figure 3 visually represents these navigation patterns.

g) Analysis of Five-Sequence Itemsets:

In the four-sequence itemset analysis, 721 rules were generated. Due to the large volume of rules, a minimum support threshold of 0.00006850 was applied to filter significant patterns. Table 4 presents key rules with high lift values. It was found that users transitioning through PSN-News (P16), PSN-Sports (P17) and Travel were likely to visit either the Front Page (P1) or the On-air (P6) page next. Figure 4 provides a graphical representation of these four-sequence itemsets.

Table 5: Pattern Discovery From Five-Sequence Itemsets

Rule No.	Rule	Highest Lift	Predicted Subsequent Page
R1	P16| P3| P11| P10 | P12	0.2071	P6
R2	P16| P3| P13 | P10 | P4	0.2071	P6
R3	P16| P3| P12| P10 | P6	0.1550	P2
R4	P17 | P3| P13 | P10 | P2	0.2071	P6
R5	P16| P5 | P14| P11 | P12	0.2071	P6
R6	P16| P5 | P13 | P6 | P1	0.1550	P2
R7	P16| P5 | P10| P9 | P12	0.2071	P6
R8	P16| P5 | P13 | P12 | P1	0.1550	P2
R9	P16| P15 | P10 | P8 | P9	0.3123	P1
R10	P17 | P15 | P8| P9 | P13	0.3123	P1

R11	P17 | P15 | P8 | P10| P9	0.3123	P1
R12	P17 | P15 | P8 | P9 | P14	0.3123	P1
R13	P17 | P15 | P8 | P9 | P2	0.3123	P1
R14	P17 | P13 | P11| P10| P9	0.3123	P1
R15	P17 | P13 | P11| P8 | P7	0.2071	P6
R16	P17 | P13 | P10 | P6| P5	0.1090	P14
R17	P17 | P13 | P10 | P6| P2	0.2071	P6
R18	P17 | P13 | P10 | P11 | P7	0.1120	P4
R19	P17 | P13 | P10 | P7 | P12	0.1120	P4
R20	P17 | P13 | P10 | P6| P4	0.1090	P14
R21	P17 | P13 | P10 | P11 | P3	0.1090	P14
R22	P17 | P13 | P8 | P7 | P12	0.2071	P6
R23	P17 | P13 | P8 | P12 | P3	0.2071	P6
R24	P17 | P13 | P8 | P3 | P6	0.1550	P2
R25	P17 | P13 | P7| P3 | P1	0.2071	P6
R26	P17 | P13 | P7 | P3 | P2	0.2071	P6
R27	P17 | P13 | P9 | P5| P3	0.3123	P1

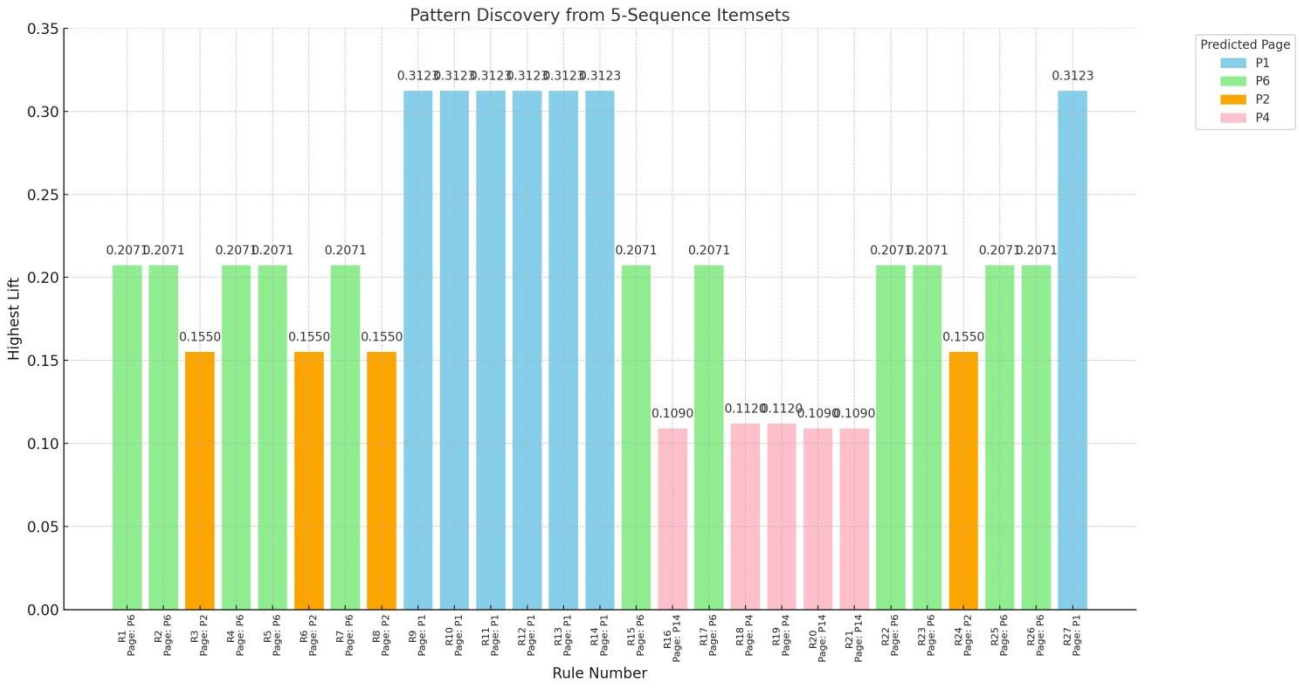

Figure 5: Prediction of Web Pages Using 5-Sequence Itemsets

h) Findings from Five-Sequence Itemsets:

In the five-sequence itemset analysis, 6304 rules were generated with a minimum support threshold of 0.00002910. The results indicate that web pages such as P6, P1, P2, P4 and P14 consistently emerge as prominent subsequent pages. Rule R1, for example, identifies P6 as the most likely next page following the pattern P16| P3| P11| P10 | P12, with a lift of 0.2071.

Similarly, pages P1 and P6 are frequently predicted to follow sequences involving PSN-News and PSN-Sports. The identified three, four and five-sequence patterns collectively illustrate the user's browsing behavior and suggest the potential pages that users are likely to visit next. These findings validate that the proposed approach can effectively predict subsequent web pages across varying sequence lengths. Figure 5 visually presents the significant patterns from five-sequence page views. Furthermore, the analysis revealed that the highest lift values were associated with P1 and P6, indicating their significance as destination pages. Pages P1, P2, P4, P6, P12 and P14 from a common group of highly visited

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 1
***Article Received:** 25 October 2020 **Revised:** 12 November 2020 **Accepted:** 30 December 2020*

pages, suggesting a strong navigational pattern among users on the website (Chifu & Salomie 2009).

v. Rule Consequent Analysis

The rule consequent analysis focused on identifying page visit trends based on web pages that appeared as

rule consequents. With a support threshold of 1.3E-05 and a confidence level of 100%, 502,384 rules in total were created. Considering the vast number of rules, only those that met the filtering criteria of a minimum support threshold of 0.00001210 and 100% confidence were selected for further analysis (Jacob et.al 2013). Each page on the website was analyzed independently to determine the number of visitors it attracted. The total number of rules associated with each page was recorded using the filtering criteria outlined earlier. The results, visualized in Figure 6, Classify the importance of web pages according to the quantity of rules that refer to them.

a) The analysis identified two distinct groups of pages based on visitor interest:

- High - interest Pages: These pages generated 3807 rules, indicating that they were frequently visited.
- Moderate - interest Pages: These pages generated 3072 rules, reflecting relatively lower engagement.

The high-interest pages included P1, P2, P4, P5, P6, P8, P10, P11, P12 and P14. In contrast, the moderate-interest pages consisted of P3, P7, P9, and P13. Notably, the rules for P15, P16, and P17 were unrelated to the 10, P11, P12, and P14. or moderate interest categories. They can, therefore, be taken to mean that these pages were essentially clicked by users, not at all or rarely.

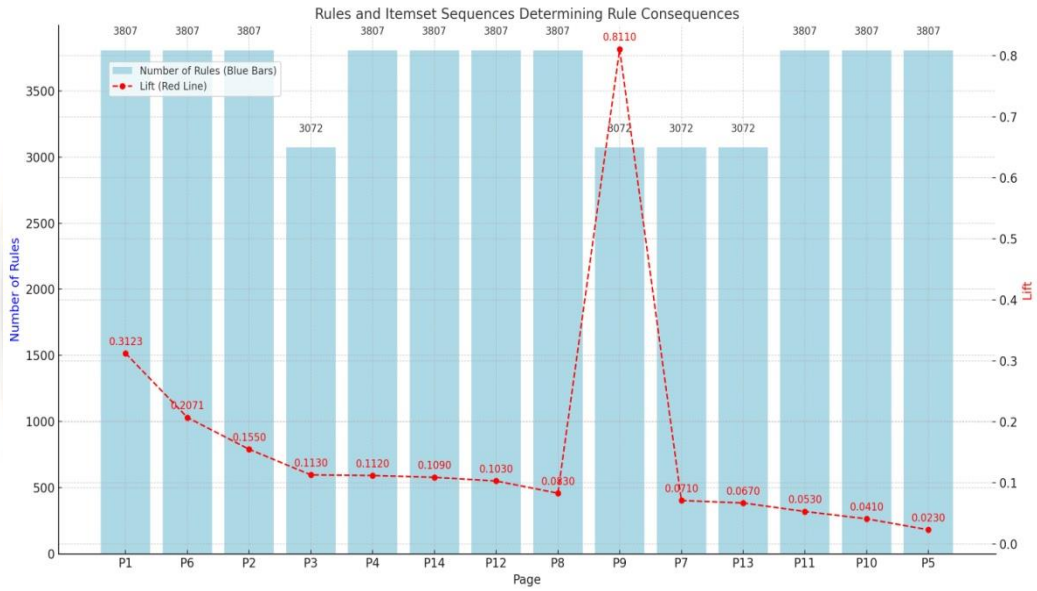

Figure 6: Based On Rule Consequences, the Number of Pages and Rules

b) Insights of the Rule Consequent Analysis

The peculiarities of web usage were analyzed with the use of the Apriori-PT algorithm's association rule mining. The analysis revealed key trends:

- Some of the pages that consistently ranked as having many visitors were P1 (Front Page) and P6 (On-Air).
- These high-interest pages relate directly to the user's browsing patterns and reveal that people are most interested in sections such as current affairs, sports and business articles.
- On the other hand, particular pages such as P3: Opinion, P7: Local, and P9: Summary clearly showed the users' less engagement or interest in those web pages.
- The lack of relevance of P15, P16 and P17 in the observed access patterns or their complete invisibility to users has been demonstrated (Kotsiantis et.al 2007).

From this work, we can realize that the application of association rule mining can reveal viable patterns in terms of Web traffic. Knowing which pages are most likely to be revisited helps businesses develop better strategies for engaging users effectively, predictions of user's web pages access behavior and improving navigation.

Conclusion

In this paper proposed web usage mining through data mining techniques for web, with specific reference to the Apriori-PT algorithm, to understand the traffic

pattern in end user's visit to web pages in providing PSNBC dataset. The first goal was to predict how the users would navigate on different web pages using string-matching functions coordinated with high-speed web interfaces. Due to the sheer size of the dataset, many rules were extracted and the vast number needed to be controlled using a minimal support level for filtering. The analysis extracted patterns from both rule antecedents and consequents to identify frequently accessed page locations and the navigation sequences of those locations. The current study showed that P1, P2, P4, P5, P6, P8, P10, P11, P12 and P14 were the most frequently accessed and the confidence co-efficient achieved was 100 percent with high Lift values. In addition, the low-interaction pages were P3, P7, P9 and P13. However, looking for the following most likely access pages, no compelling rules were applied to pages P15, P16 and P17, suggesting that they had very little or no hits by the users. The study shows that association rule mining helps identify more significant usage patterns from big data. This approach directs the context for user's perceived preferences, enabling organizations to adjust the content on their websites and pathways to efficiently attract their user's attention. This study's future work can be extended as follows: Using dynamic prediction models and incorporating additional datasets for comparison.

References

1. Agrawal, R. & Srikant, R., 1994. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 1215(1), pp.487–499.
2. Anitha, A. & Krishnan, N., 2011. A dynamic web mining framework for e-learning recommendations using rough sets and association rule mining. International Journal of Computer Applications, 12(11), pp.19–25.
3. Babu, K.G., Komali, A., Mythry, V. & Ratnam, A.S.K., 2000. Web mining using semantic data mining techniques. International Journal of Soft Computing Engineering (IJSCE), 3(2), pp.2231–2307.
4. Chakrabarti, S., 2002. Mining the Web: Analysis of hypertext and semi-structured data. Morgan Kaufmann.
5. Chandra, B. & Basker, S., 2000. A new approach for classification of patterns having categorical attributes. IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp.960–964.
6. Chifu, V. & Salomie, I., 2009. A fluent calculus approach to automatic web service composition. Advances in Electrical and Computer Engineering, 9(3), pp.75–83.
7. Chun-sheng, Z. & Li, Y., 2014. Extension of local association rules mining algorithm based on Apriori algorithm, pp.340–343.
8. Debahuti, M., 2010. Predictive data mining: Promising future and applications. International Journal of Computer and Communication Technology, 2(1), pp.20–28.
9. Eirinaki, M., Vazirgiannis, M. & Kapogiannis, D., 2005. Web path recommendations based on page ranking and Markov models. Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management, pp.2–9.
10. Ganapathy, S., Sethukkarasi, R., Yogesh, P., Vijayakumar, R. & Kannan, A., 2014. An intelligent temporal pattern classification system using fuzzy temporal rules and particle swarm optimization. Sadhana - Academy Proceedings in Engineering Sciences, 39(2), pp.283–302.
11. Gao, S., Alhajj, R., Rokne, J. & Guan, J., 2009. Set-based approach in mining sequential patterns. IEEE 24th International Symposium on Computer and Information Sciences (ISCIS 2009), pp.218–223.
12. Hacibeyoglu, M., Arslan, S. & Kahramanli, S., 2013. A hybrid method for fast finding the reduct with the best classification accuracy. Advances in Electrical and Computer Engineering, 13(4), pp.57–64.
13. Han, J. & Kamber, M., 2011. Data Mining – Concepts and Techniques. 3rd ed. Morgan Kaufmann Publishers.
14. Hung, Y.S., Chen, K.L.B., Yang, C.T. & Deng, G.F., 2013. Web usage mining for analysing elder self-care behavior patterns. Expert Systems with Applications, 40(2), pp.775–783.
15. Jacob, S.G. & Ramani, R.G., 2012. Evolving efficient classification rules from cardiotocography data through data mining methods and techniques. European Journal of Scientific Research, 78(3), pp.468–480.
16. Jacob, S.G. & Ramani, R.G., 2013. Design and implementation of a clinical data classifier: A supervised learning approach. Research Journal of Biotechnology, 8(2), pp.16–24.
17. Jacob, S.G., Ramani, R.G. & Nancy, P., 2013. Discovery of knowledge patterns in lymphographic clinical data through data mining methods and techniques. Advances in Computing and Information Technology. LNCS Springer, pp.129–140.
18. Kaur, J., 2015. Association rule mining: A survey. International Journal of Hybrid Information Technology, 8(7), pp.239–242.
19. Koo, B.B., 2016. Association rule mining and genetic algorithm (GA) for data mining-based intrusion detection system: A review approach. IEEE Journal on Selected Areas in Communications, 34(3), pp.1–6.
20. Kotsiantis, S.B. & Kanellopoulos, D., 2001. Association rules mining: A recent overview. GESTS International Transactions on Computer Science and Engineering, 32(1), pp.71–82.

21. Kotsiantis, S.B., Zaharakis, I.D. & Pintelas, P.E., 2007. Supervised machine learning: A review of classification techniques, pp.3–24.
22. Kriegel, H.P., 2007. Future trends in data mining. Data Mining Knowledge Discovery, 15(1), pp.87–97.
23. Kum, H.-C., Paulsen, S. & Wang, W., 2005. Comparative study of sequential pattern mining frameworks: Support framework vs. multiple alignment framework. IEEE 2nd International Conference on Data Mining - Workshop on the Foundation of Data Mining and Discovery, pp.43–70.
24. Kumar, S.K. & Chezian, R.M., 2012. A survey on association rule mining using Apriori algorithm. International Journal of Computer Applications, 45(5), pp.7–50.
25. Mgiba, F. M. (2020). Artificial intelligence, marketing management, and ethics: their effect on customer loyalty intentions: a conceptual study. The Retail and Marketing Review, 16(2), 18-35.
26. Mitić, V. (2019). Benefits of artificial intelligence and machine learning in marketing. In Sinteza 2019-International scientific conference on information technology and data related research (pp. 472-477). Signinum University
27. Patil, A. & Gupta, P., 2017. A review of the up-growth algorithm using association rule mining. 2017 International Conference on Computing Methodologies and Communication (ICCMC). DOI: 10.1109/iccmc.2017.8282605.
28. Prithiviraj, P. & Porkodi, R., 2015. A comparative analysis of association rule mining algorithms in data mining: A study. American Journal of Computer Science and Engineering Survey (AJCSES), 3(1), pp.98–119.
29. Rahman, T., Kabir, M.M.J. & Kabir, M., 2019. Performance evaluation of fuzzy association rule mining algorithms. 2019 4th International Conference on Electrical Information and Communication Technology (EICT).
30. Srinivasan, J., Cooley, R., Deshpande, M. & Tan, P.N., 2000. Web usage mining: Discovery and applications of usage patterns from web data. ACM SIGKDD Explorations Newsletter, 1(2), pp.12–23.
31. Zanker, M., Rook, L., & Jannach, D. (2019). Measuring the impact of online personalisation: Past, present and future. International Journal of Human-Computer Studies, 131. https://doi.org/10.1016/j.ijhcs.2019.06.006.