ANALYSE USER BEHAVIOUR & OPTIMISE THE USER WORKFLOW USING A MACHINE LEARNING ALGORITHM

Introduction:

The adoption of data science and machine learning for marketing has been on the rise from small to large organizations

E- commerce has resulted in organizations investing significant resources in online strategies to extend business processes on to the World Wide Web. Traditional methods of measuring Web usage fall short of the richness of data required for the effective evaluation of such strategies. Web analytics are an approach that may meet organizational demand for effective evaluation of online strategies.

A website user is a person who is accessing, browsing or interacting with a website, and user behavior refers to how people use a website. Behaviors include everything from the journey they take through the site to interactions such as clicks. When it comes to optimizing a website, simply monitoring behavior can only get you so far. The real value comes from analyzing users' actions to get to the bottom of what makes them behave as they do.

Behaviour is complex and varies across different websites depending on the target audience. This means you need to learn specifically about your users. Who are they? What are their needs? Which browsers and devices do they prefer? How often do they purchase? Answering these questions is crucial if you want to have a competitive edge, meet consumer needs and retain your customers. By researching online behaviours, you can get an idea of what users are trying to achieve, the factors driving certain behaviours, where they experience friction and areas where user experience can be better. Ultimately, learning how visitors behave on your website allows you to provide an enhanced experience that's in line with user needs, which in turn will ensure your business continues to grow.

One of the main tools we use to carry out research is Google Analytics. This is free and fairly easy to set up. Once installed, you'll have access to valuable data about how users behave on your site, including where they land and go next, where they drop off and what they interact with. You can also use it to discover overall trends and patterns and source opportunities for

Web Application

E-commerce websites are online portals that facilitate online transactions of goods and services through means of the transfer of information and funds over the Internet. In the early days, e-Commerce was done partially through emails and phone calls. Now, with a single website, anything and everything that a transaction needs, can be executed online.

I have used following steps to create my E-commerce website:

- Firstly, I have taken domain name(sachinmundhe.co) (website-SACHIN MUNDHE – Substantial world) and hosting from GoDaddy.com, which also provided me access to the wordpress.
- Then, I use a pre-existent theme Astra from wordpress and edit the same to suit my requirements.

- My website contains following webpages:

1) **Blog**-companies details

2) **Checkout**-Whatever Customers added to basket

3) **Shop** - It contains team details.

4) **Home**- It is the homepage of my website.

5) **Look Book**-Overview of product

6) **Store**- It contains various products categorized into different categories.

7) **My Account**- It contains account details of the user.

8) **Contact** - It contains the contact details of the company.

- Website contains various features:

a) View a list of products

b) View product details

c) Search products

d) Use filters to sort the product (Popularity, Rating, Price(High-Low)wise)

e) Add a product to the Basket.

f) Apply Coupon.

g) Proceed to Checkout.

h) Payment Gateway.

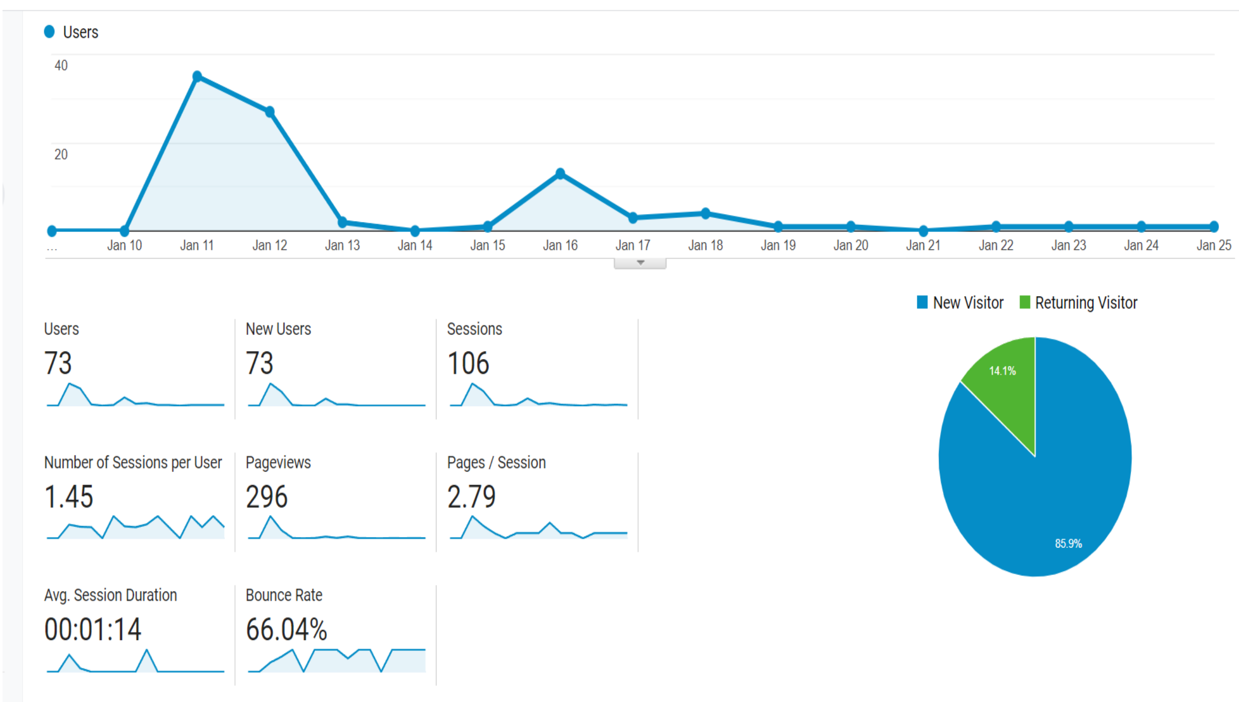
i) User can give rating the product and can put their Name and email for getting acknowledgment feedback

Analytics

User Behavior Analytics or UBA focuses on what the user is doing: apps launched, network activity, and, most critically files accessed (when the file or email was touched, who touched it, what was done with it and how frequently). UBA technology searches for patterns of usage that indicate unusual or anomalous behavior — regardless of whether the activities are coming from a hacker, insider, or even malware or other processes. While UBA won't prevent hackers or insiders from getting into your system, it can quickly spot their work and minimize damage. I have integrated my website to analytics platform named Google Analytics. This tool is able to track the users action and the time spent by the users on the various actions of the website. I have accumulated the analytics data from various users and stored it.

I have collected data of around 71 users with various data points such as:

- Time spent by users on the different pages on website
- Filters being used by the users
- Time spent by users on particular product



Audience Overview

	Page	Pageviews	Unique Pageviews	Avg. Time on Page	Entrances	Bounce Rate	% Exit	Page Value
		318 % of Total: 100.00% (318)	193 % of Total: 100.00% (193)	00:00:46 Avg for View: 00:00:46 (0.00%)	109 % of Total: 100.00% (109)	64.22% Avg for View: 64.22% (0.00%)	34.28% Avg for View: 34.28% (0.00%)	₹0.00 % of Total: 0.00% (₹0.00)
1.	/	133 (41.82%)	96 (49.74%)	00:01:11	95 (87.16%)	62.11%	61.65%	₹0.00 (0.00%)
2.	/shop/	84 (26.42%)	37 (19.17%)	00:00:35	9 (8.26%)	77.78%	15.48%	₹0.00 (0.00%)
3.	/my-cart/	27 (8.49%)	5 (2.59%)	00:00:43	1 (0.92%)	0.00%	3.70%	₹0.00 (0.00%)
4.	/shop/page/2/	23 (7.23%)	15 (7.77%)	00:00:37	0 (0.00%)	0.00%	8.70%	₹0.00 (0.00%)
5.	/product/home-lamp/	9 (2.83%)	3 (1.55%)	00:00:14	0 (0.00%)	0.00%	0.00%	₹0.00 (0.00%)
6.	/blog/	8 (2.52%)	5 (2.59%)	00:01:20	0 (0.00%)	0.00%	12.50%	₹0.00 (0.00%)
7.	/product/grey-pulllover/	7 (2.20%)	6 (3.11%)	00:00:19	0 (0.00%)	0.00%	0.00%	₹0.00 (0.00%)
8.	/product/product-six/	6 (1.89%)	6 (3.11%)	00:00:22	1 (0.92%)	100.00%	50.00%	₹0.00 (0.00%)
9.	/checkout/	5 (1.57%)	4 (2.07%)	00:01:51	1 (0.92%)	100.00%	40.00%	₹0.00 (0.00%)
10.	/look-book/	4 (1.26%)	4 (2.07%)	00:00:26	0 (0.00%)	0.00%	25.00%	₹0.00 (0.00%)
11.	/contact/	3 (0.94%)	3 (1.55%)	00:00:17	0 (0.00%)	0.00%	0.00%	₹0.00 (0.00%)
12.	/my-account/	3 (0.94%)	3 (1.55%)	00:00:00	1 (0.92%)	100.00%	100.00%	₹0.00 (0.00%)
13.	/product/mens-red-short/	2 (0.63%)	2 (1.04%)	00:00:09	0 (0.00%)	0.00%	0.00%	₹0.00 (0.00%)
14.	/product/mens-yellow-shirt/	2 (0.63%)	2 (1.04%)	00:00:04	1 (0.92%)	100.00%	50.00%	₹0.00 (0.00%)
15.	/blog-post-title/	1 (0.31%)	1 (0.52%)	00:03:12	0 (0.00%)	0.00%	0.00%	₹0.00 (0.00%)
16.	/product/mens-black-shirt/	1 (0.31%)	1 (0.52%)	00:00:21	0 (0.00%)	0.00%	0.00%	₹0.00 (0.00%)

Example – Analytics All Website Data Pages

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	A1	A1	A1	A1	A2	A1	A1	A1	A1	A1	A3	A4	A3	A5	A6	A4	A7	A4	A7	A4	A7	A3	A4	A8	A4	A3	A4	A3	A1	A4	A3	A1	A1
2	A9	A9	A1	A1	A1	A1	A1	A10	A10	A3	A10	A10	A3																				
3	A8	A1	A4	A3	A7	A7	A4	A7	A4	A8	A4	A7	A7	A3	A11	A4	A11	A3	A4	A4	A1	A8	A4	A3	A1	A1							
4	A1	A1																															
5	A1	A1	A1																														
6	A1	A1																															
7	A1	A1	A12	A3	A11	A3	A1																										
8	A1	A1																															
9	A12	A12	A3	A1	A13	A5	A3	A10	A3	A1																							
10	A1	A1																															
11	A1	A10	A3	A1	A1	A3	A11	A3	A1																								
12	A1																																
13	A1																																
14	A9	A1																															
15	A1																																
16	A1																																
17	A1	A3	A10	A3	A1																												
18	A8	A4	A4	A3	A4	A3	A4	A11	A3	A4	A3	A1																					
19	A1																																
20	A1																																
21	A1	A3	A10	A3	A1																												
22	A1																																
23	A1																																
24	A1	A3	A1																														
25	A1	A3	A1	A3	A10	A3	A10	A3	A1																								

Example- Action Perform By Users

ML Model and Data Analysis

The elusive clickstream data. Many platforms, like Facebook rely on these generated data from what a user clicks and what doesn't. To start analyzing clickstream data, we need first to be able to capture step by step a user's activity across a web page or application. And that is of great value in the hands of any Internet marketer. Getting a 360-degree view of a customer by knowing what he is clicking and what he is not can get you a huge improvement in both your products and your customers' experience

Data Collection

Either you have your data in your data warehouse, or you need to enrich it with more data sources you need to have a way to collect and store data consistently into a database.

Data Preparation

Raw data is like a rough diamond; It requires some refinement before being truly valuable. In the data world, refinement includes data processing, cleaning, and transformation of the initial data into something convenient for the analysis you are going to carry out. In this case, we would like to have our data grouped into users. It would be good too, we could arrange the events of each user in time order before moving to actual analysis. In contrast to other data sequences, clickstream data can have varying length for every different users.

In order to transform the initially collected event log into clickstream data we need to:

- Identify events/actions performed by the same user and group them together
- Split them further into subgroups of events based on which of those were performed during the same session according to the session's definition given above.

At this point the dataset we are going to use for the rest of the analysis should look like this

49	49	A1														
50	50	A1	A3	A10	A3	A1										
51	51	A1	A3	A10	A3	A2	A3	A14	A3	A15	A3	A11	A3	A1	A13	A1
52	52	A1														
53	53	A1	A3	A1	A3	A1	A3	A10	A3	A1						
54	54	A1														
55	55	A1														
56	56	A3	A1													
57	57	A1														
58	58	A1														
59	59	A1														
60	60	A1	A3	A12	A3	A7	A3	A1	A3	A1						
61	61	A3	A14	A3	A3	A1										
62	62	A1	A3	A10	A3	A1										
63	63	A1														
64	64	A1														
65	65	A1														
66	66	A1	A1													
67	67	A1														
68	68	A3	A1													
69	69	A1														
70	70	A13	A13	A1												
71	71	A1														

In this representation, each line corresponds to a user. The first field is the user's name while the next fields the actions performed by the user during this session

Model Construction

As in most cases, the methods we can deploy for solving this problem are many

Markov Chains

A stochastic process is called a Markov process if the state of the random variable at the next instance of time depends only on the outcome of the random variable at the current time. A Markov chain is, in fact, a Markov process too in either discrete or continuous time with a countable state space.

In clickstream analysis, we usually utilize these Markov Chains. The process takes the state from a finite set at each time . The order of a Markov Chain is derived from the number of recent states on which the current state, we assume, depends. Based on this, zero-order chains imply that the probability of being in a state in the next step is independent of all previous states.

Higher order Markov Chain introduced by the Raftery (1985) will lead to more realistic models. At the same time, the parameters needed for the representation increase exponentially and so it is important to find a right balance between these two

Fitting a Markov Chain

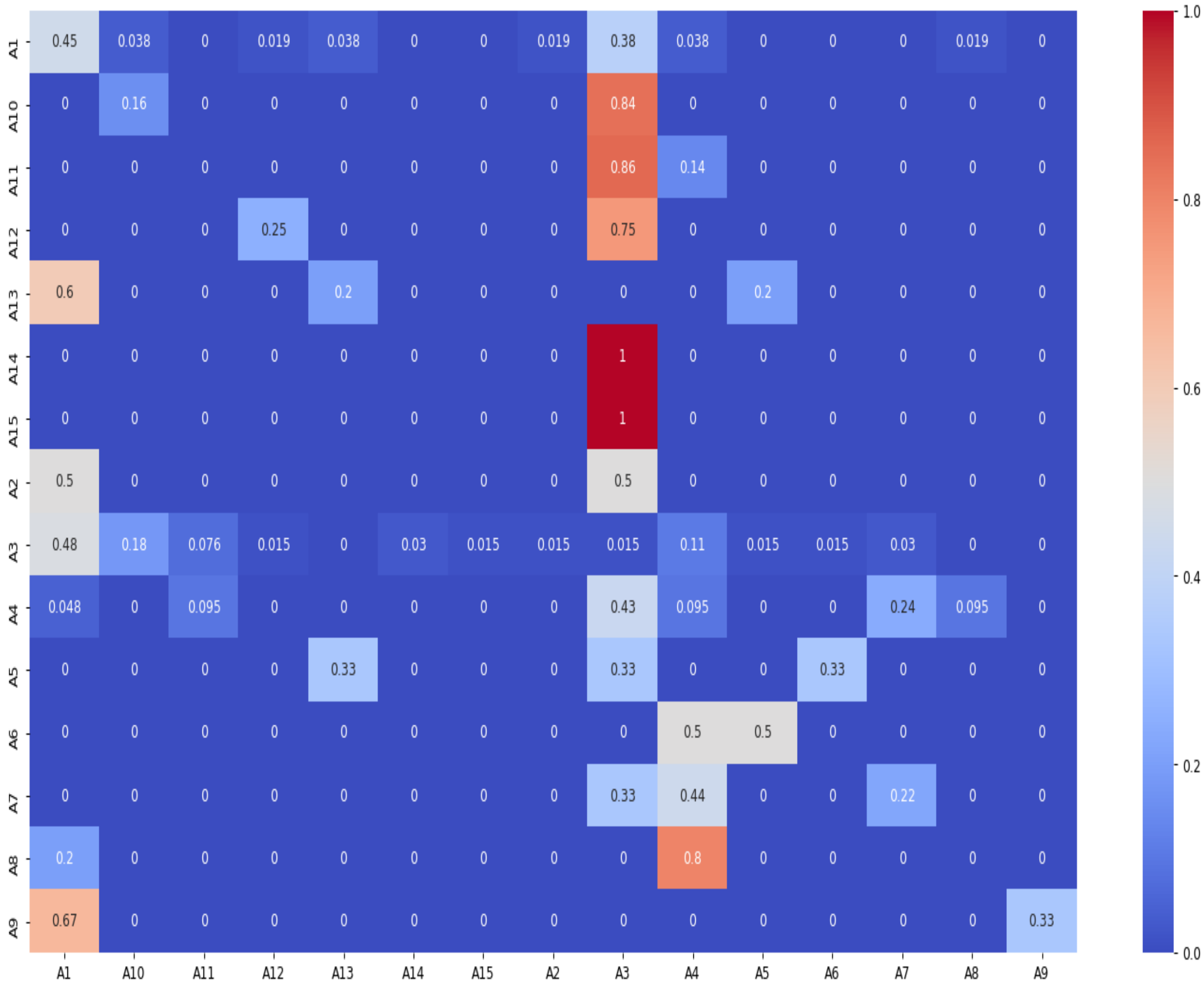
As mentioned before at this point our dataset looks like

49	49	A1														
50	50	A1	A3	A10	A3	A1										
51	51	A1	A3	A10	A3	A2	A3	A14	A3	A15	A3	A11	A3	A1	A13	A1
52	52	A1														
53	53	A1	A3	A1	A3	A1	A3	A10	A3	A1						
54	54	A1														
55	55	A1														
56	56	A3	A1													
57	57	A1														
58	58	A1														
59	59	A1														
60	60	A1	A3	A12	A3	A7	A3	A1	A3	A1						
61	61	A3	A14	A3	A3	A1										
62	62	A1	A3	A10	A3	A1										
63	63	A1														
64	64	A1														
65	65	A1														
66	66	A1	A1													
67	67	A1														
68	68	A3	A1													
69	69	A1														
70	70	A13	A13	A1												
71	71	A1														

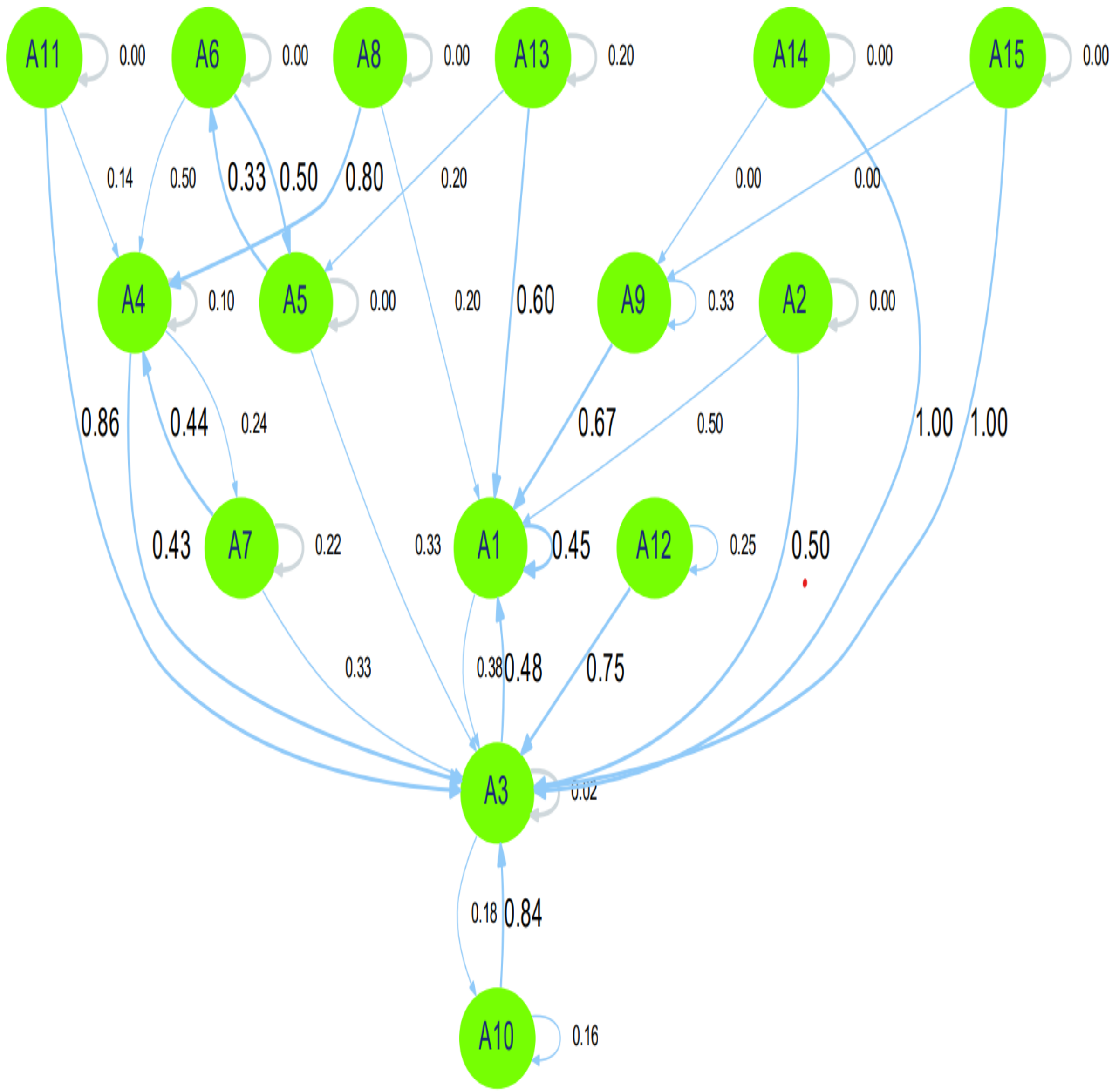
Fitting the Markov Chain model gives us the transition probabilities matrices and the lambda parameters of the chain for each one of the three lags along with the Start and End Probabilities.

Start and End probabilities correspond to the probability that a clickstream will start or end with this specific event. The transition probability matrix can be represented as a heat map with the y-axis representing the current state and x-axis the next one. The more reddish the color, the more probable the indicated transition will occur.

Heat Map:


A1 0.45 0.038 0 0.019 0.038 0 0 0.019 0.38 0.038 0 0 0 0.019 0
A10 0 0.16 0 0 0 0 0 0 0.84 0 0 0 0 0 0
A11 0 0 0 0 0 0 0 0 0.86 0.14 0 0 0 0 0
A12 0 0 0 0.25 0 0 0 0 0.75 0 0 0 0 0 0
A13 0.6 0 0 0 0.2 0 0 0 0 0 0.2 0 0 0 0
A14 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
A15 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
A2 0.5 0 0 0 0 0 0 0 0.5 0 0 0 0 0 0
A3 0.48 0.18 0.076 0.015 0 0.03 0.015 0.015 0.015 0.11 0.015 0.015 0.03 0 0
A4 0.048 0 0.095 0 0 0 0 0 0.43 0.095 0 0 0.24 0.095 0
A5 0 0 0 0 0.33 0 0 0 0.33 0 0 0.33 0 0 0
A6 0 0 0 0 0 0 0 0 0 0.5 0.5 0 0 0 0
A7 0 0 0 0 0 0 0 0 0.33 0.44 0 0 0.22 0 0
A8 0.2 0 0 0 0 0 0 0 0 0.8 0 0 0 0 0
A9 0.67 0 0 0 0 0 0 0 0 0 0 0 0 0 0.33
A1 A10 A11 A12 A13 A14 A15 A2 A3 A4 A5 A6 A7 A8 A9
1.0
0.8
0.6
0.4
0.2
0.0

Transaction Diagram:



ANALYSE USER BEHAVIOUR & OPTIMISE THE USER WORKFLOW USING A MACHINE LEARNING ALGORITHM

Actions Details

A1 - Home

A2 - product mens-yellow-shirt

A3 - shop

A4 - my-cart

A5 - contact

A6 - look-book

A7 - product/home-lamp

A8 - checkout

A9 - my-account

A10 - /shop/page/2/

A11 - /product/grey-pulllover/

A12 - /product/product-six/

A13 - blog

A14 - /product/mens-red-short/

A15 - /product/mens-black-shirt/

Suggestion to UX Designer

- The following are the insights from the Heat Map generated by the Markov chain model:
- The transaction from Action A15 to Action 3, Action 14 to Action A3, and Action A11 to Action A3 is more co-related to each other.
- The transaction from Action A1 to A9 is not corelated to each other

Suggestion taken from transaction diagram along with the corresponding probability are:

- All the users must start from A1 (homepage).
- The user from action A2, A9, A12, A14, A14, A15 (all are products) are occasionally not going to A4(cart). They should go to A4 (cart) so the UX designer should guide the user to go to the A4 from these actions.
- The UX designer should guide the user to go to the A9 (My Account) from A1 (Home).
- Users are either going back to A4 (cart) or close the application (Checkout) without proceeding to payment, so the UX designer should guide the user to proceed to checkout and make payment.
- FromA1 (Home page)UX designer should inform user about sorting options.
- From A1 (homepage)UX designer should guide the user to go to next actions A8(checkout)page for billing details for placing an order.