

```
1 <body oninput=javascript:alert(1)><input autofocus>
2 <math href="javascript:javascript:alert(1)">CLICKME</math> <math> <maction actiontype="statusline#http://google.com">
3 <frameset onload=javascript:alert(1)>
4 <table background="javascript:javascript:alert(1)">
5 <!--
6 <comment>
7 <![>
8 <style>
9 <li style=list-style:url() onerror=javascript:alert(1)> <div style=content:url(data:image/svg+xml,%%3Csvg%%3E
10 <head><base href="javascript://"></head><body><a href="/. /, javascript:alert(1)//#">XXX</a></body>
11 <SCRIPT FOR=document EVENT=onreadystatechange>javascript:alert(1)</SCRIPT>
12 <OBJECT CLASSID="clsid:333C7BC4-460F-11D0-BC04-0080C7055A83"><PARAM NAME="DataURL" VALUE="javascript:alert(1)"
13 <object data="data:text/html;base64,%(base64)s">
```

A	B	C	D	E	F	G	H	I	J
	Method	User-Agent	Pragma	Cache-Control	Accept	Accept-encoding	Accept-char	language	host
Normal	GET	Mozilla/5.0	no-cache	no-cache	text/xml,application/xml,application/xhtml+xml,text/css;q=0.7,*/*;q=0.6	x-gzip, x-deflate, gzip, deflate	utf-8, utf-8; en		localhost:8 JSESSIONID=9E0C8F0A40D989B37929216000000000
Normal	GET	Mozilla/5.0	no-cache	no-cache	text/xml,application/xml,application/xhtml+xml;text/css;q=0.7,*/*;q=0.6	x-gzip, x-deflate, gzip, deflate	utf-8, utf-8; en		localhost:8 JSESSIONID=9E0C8F0A40D989B37929216000000000
Normal	POST	Mozilla/5.0	no-cache	no-cache	text/xml,application/xml,application/xhtml+xml;text/css;q=0.7,*/*;q=0.6	x-gzip, x-deflate, gzip, deflate	utf-8, utf-8; en		localhost:8 JSESSIONID=9E0C8F0A40D989B37929216000000000
Normal	GET	Mozilla/5.0	no-cache	no-cache	text/xml,application/xml,application/xhtml+xml;text/css;q=0.7,*/*;q=0.6	x-gzip, x-deflate, gzip, deflate	utf-8, utf-8; en		localhost:8 JSESSIONID=9E0C8F0A40D989B37929216000000000
Normal	POST	Mozilla/5.0	no-cache	no-cache	text/xml,application/xml,application/xhtml+xml;text/css;q=0.7,*/*;q=0.6	x-gzip, x-deflate, gzip, deflate	utf-8, utf-8; en		localhost:8 JSESSIONID=9E0C8F0A40D989B37929216000000000
Normal	GET	Mozilla/5.0	no-cache	no-cache	text/xml,application/xml,application/xhtml+xml;text/css;q=0.7,*/*;q=0.6	x-gzip, x-deflate, gzip, deflate	utf-8, utf-8; en		localhost:8 JSESSIONID=9E0C8F0A40D989B37929216000000000

	B	C	D
	Sentence	Label	
0	<li><a href="/wiki/File:Socrates.png" class="image"><im	0	
1	<tt onmouseover="alert(1)">test</tt>	1	
2	+ +⊕⊕1⊕+ +⊕⊕1⊕ -⊕1⊕⊕⊕++⊕+	0	
3	+ +⊕⊕1⊕+ +⊕⊕1⊕ -⊕1⊕⊕⊕++⊕+	0	
4	+ +⊕⊕1⊕++ ++ ⊕⊕↑⊕   ⊕⊕⊕⊕	0	
5	<li id="cite_note-118"><span class="mw-cite-backlink">•	0	
6	<li><a href="/wiki/Contextualism" title="Contextualism">(	0	
7	<li id="cite_note-Representing_causation-95"><span cla	0	
8	<tr><td class="plainlist" style="padding:0 0.1em 0.4em">	0	
9	+ +⊕⊕1⊕+	0	

4. <https://github.com/fmreani/Cross-Site-Scripting-XSS> this link has three dataset named Payload.csv, XSSTesting.csv and XSSTraining.csv. In payload.csv 2 parameter found like below

A	
Payloads	Class
http://www.nwce.gov.uk/search_process.php?keyword=%22%3e%3cscript%3ealert	Malicious
http://www.manchester.gov.uk/site/scripts/google_results.php?q=%22%3e%253csc	Malicious
http://www.ldsmissions.com/us/index.php?action=missionary.info%3cmarquee%3ep	Malicious
http://education.powys.gov.uk/english/adult_ed/register.php?lforenam=%22%3e%3c	Malicious
http://www.northwarks.gov.uk/site/scripts/google_results.php?q=%22%3e%253csc	Malicious
http://www.chaoticwars.co.uk/register.php?ref='%3e%3cifframe%20src=http://google	Malicious

and Others has same parameters which is used 0 and 1 to detect various special character and input in URL and then detect the url is Malicious or Bening

Contains &lt	ScripTag	Readable	Contains ><	Contains '><	Contains And	Contains Percentage	Contains Slash	Contains BackSlash
-----------------	----------	----------	----------------	-----------------	-----------------	------------------------	-------------------	-----------------------

AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN
is C	Contains C	Contains t	Contains S	Contains C	Contains C	Contains C	Contains C	Contains V	Contains F	Contains E	Contains A	Contains E	Letters Ra	Numbers	Symbols R	Class
0	0	0	0	1	0	0	1	0	0	0	1	0	0.7614	0.0114	0.2273	Malicious
0	0	0	0	1	0	0	1	0	0	0	1	0	0.7614	0.0114	0.2273	Malicious
0	0	0	0	1	0	0	1	0	0	0	1	0	0.7203	0.0593	0.2203	Malicious
0	0	0	0	0	0	0	1	0	0	0	1	0	0.7071	0.1065	0.1864	Malicious
0	0	0	0	0	0	0	1	0	0	0	1	0	0.7158	0.1053	0.1789	Malicious
0	0	0	0	0	0	0	1	0	0	0	0	0	0.6483	0.1517	0.2	Malicious

5. <https://github.com/fawaz2015/XSS-dataset> another url analyzer type dataset which has 167 parameters to understand input URL

<ul style="list-style-type: none"> <li>'url_special_characters',</li> <li>'url_tag_script',</li> <li>'url_tag_iframe',</li> <li>'url_tag_meta',</li> <li>'url_tag_applet',</li> <li>'url_tag_object',</li> <li>'url_tag_embed',</li> <li>'url_tag_link',</li> <li>'url_tag_svg',</li> <li>'url_tag_frame',</li> <li>'url_tag_form',</li> <li>'url_tag_div',</li> <li>'url_tag_style',</li> <li>'url_tag_video',</li> <li>'url_tag_img',</li> <li>'url_tag_input',</li> <li>'url_tag_textarea',</li> <li>'url_attr_action',</li> <li>'url_attr_archive',</li> <li>'url_attr_background',</li> <li>'url_attr_classid',</li> </ul>	<ul style="list-style-type: none"> <li>'html_attr_background',</li> <li>'html_attr_cite',</li> <li>'html_attr_classid',</li> <li>'html_attr_codebase',</li> <li>'html_attr_data',</li> <li>'html_attr_dsync',</li> <li>'html_attr_formaction',</li> <li>'html_attr_href',</li> <li>'html_attr_icon',</li> <li>'html_attr_longdesc',</li> <li>'js_dom_windows',</li> <li>'js_dom_location',</li> <li>'js_dom_document',</li> <li>'js_prop_cookie',</li> <li>'js_prop_document',</li> <li>'js_prop_referrer',</li> <li>'js_method_write',</li> <li>'js_method_getElementById',</li> <li>'js_method_alert',</li> <li>'js_method_eval',</li> <li>'js_method_fromCharCode',</li> <li>'js_method_prompt',</li> </ul>
---	--

	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ
d	js_method	js_method	js_method	js_method	js_min_len	js_min_def	js_min_fun	js_string_r	html_length	Label	
1	0	0	0	0	11	0	1	353	11380	0	
1	0	0	0	0	267	0	0	4422	56223	0	
0	1	0	0	1	13	0	1	275	10712	1	
0	0	0	0	0	43	0	0	346	11922	0	
0	8	0	0	0	1879	2	15	1879	14735	1	
0	0	0	0	0	173	1	3	173	12324	0	
2	0	0	0	0	14	0	1	334	28684	0	
0	0	0	1	0	30	0	0	7533	151436	0	
0	1	0	0	0	10	0	0	139	15286	1	
2	0	0	0	0	82	0	2	2516	35728	0	