



ABUSE RELATED POSTS REPORTER USING WEB CRAWLER IN PYTHON

Dr. P .A. Tijare¹, Dishil Purswani², Tanishk Rathi³, Swapnil Chavhan⁴, Vijay Jangid⁵, Shubham Agrawal⁶

¹Computer Science and Engineering, Sipna College of engineering and technology Amravati

²Computer Science and Engineering, Sipna College of engineering and technology Amravati

³Computer Science and Engineering, Sipna College of engineering and technology Amravati

⁴Computer Science and Engineering, Sipna College of engineering and technology Amravati

⁵Computer Science and Engineering, Sipna College of engineering and technology Amravati

⁶Computer Science and Engineering, Sipna College of engineering and technology Amravati

Article DOI: <https://doi.org/10.36713/epra10288>

DOI No: 10.36713/epra10288

ABSTRACT

This research paper shows one of the application areas where Web Scraping can be used to solve some social issues. According to [4], Web Scraping automates this process and helps you display all the data at the same place with ease and in a small amount of time. Main objective of Web Scraping is to extract information from one or many websites and process it into simple structures such as spreadsheets, database or CSV file. In our website, we have made use of this technology to solve a important social issue. Many cases of abuse have been increased with the use of Internet. Our website allows the authorized communities to search for such abuse related posts from different websites, while staying on a single webpage. The user will just have to enter the keyword, for e.g. – women abuse, and he will get the posts from several number of websites on a single webpage. The method explained in this paper is applied to a popular social platform – ‘Twitter’ as a use case. It can be applied to a number of websites. This will help the authorities to save time in searching for such crimes and give more time to solve these issues.

KEYWORDS : *Web Scraping, Django, HTML, CSS, Python, Selenium.*

I. INTRODUCTION

Technology developments are increasing rapidly. Many information is circulating on the internet. Nowadays, all information could be search easily. Social media has recently become part of people's daily activities; many of them spend hours each day on Twitter, Messenger, Instagram, Facebook, and other popular social media. It is not only providing the information itself, but also provide the information by users' social network. The information relevancy produced by social media application is at medium-high rate. It's a plus point for information spreading using social media. Now, there have been many kinds of abuse cases reported every year on this platform.

Some of them are: - child abuse, women abuse, etc. Many times, victims do not report to the police station or concerning authorities directly. And, abuse is not limited only

to humans, it is also done with animals, which cannot report it to anyone. So, nothing is done to solve these issues. So, we got an idea to create a platform where such cases can be reported to the concerned authorities. So that these issues can be solved and more and more victims can get justice. The idea is to create a website where a keyword can be entered and then the data scraped from several websites (where the data related to that keyword can be obtained) and will provide it on the same single webpage. So that, the authorities will not have to spend time searching over different sites to find such reports.

Social media websites can be considered best for such cases as many people nowadays use hashtags to report such cases, in a hope that the concerned authorities will check these posts. So, our website has used such two main platforms, i.e., Instagram and Twitter. A web scraper is a piece of software that automates the time-consuming process of extracting valuable information from third-party websites. Typically, this



method involves sending a request to a particular web page, reading the HTML code, and sending it to the user.

II. LITERATURE REVIEW

As the Internet sets out to connect the world, more people than ever are using the Internet and the World Wide Web in their everyday lives. According to [1], estimates have found that in 2000, there were around 400 million users of the Internet, while in 2015, there were around 3.2 billion users, an increase of 2.8 billion in 15 years (ITU, 2015) along with the proportions of the populations of developed, developing and least developed countries (LDCs) who have access to the Internet. As the Internet and the World Wide Web continue to grow and amass more users, there is a higher risk of criminal activity. Currently, the web is uncontrolled and unregulated, which allows anyone to utilise it for communication and create their own websites with documents, links and their own content (Albert et al., 1999). This is, in part, due to the lack of centralised control on the web, with no governing bodies to regulate content (Castillo, 2004).

What this means is that it is possible for someone with intent to commit a crime to use an existing website or create a website to use as a platform for criminal activities. There are many types of crime committed on the Internet, ranging from viruses and malware to fraud and social media abuse.

Various research are done and predictive models are made for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is used for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps the crime analysis to analysis these crime networks by means of various interactive visualization. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. The data privacy, reliability, accuracy can be improved for enhanced prediction. It can be seen that the most common medium used to facilitate crimes is social media, with 18,712 victims in 2016 (IC3, 2017). What this shows is that criminals would often use communication tools to commit offences. One such offence is abuse. There have been numerous estimates of the number of victims of abuse in the world. Exact data on violence, abuse and exploitation is not sufficient, but overall, the nation is becoming increasingly aware of violence against children, especially sexual abuse. Several cases that may have earlier gone unnoticed, are now. Even after taking so many steps, many people are afraid of reporting and complaining to the concerned authorities. No action can be taken until an abuse is reported. Above two abuses are the ones which are paid most attention. But there are many other abuses. Consider animal abuse. Animals cannot report by themselves any harm made to them. In all such cases, web scrawling application has been developed for tackling various crime except for social media platform. So, to help the victims, many socially active people take an initiative by reporting these things on popular social platforms with the use of hashtags. This application helps in extracting all such abuse related posts from different websites

and providing it in the same place so that the authorities can get informed sooner and then they can take further actions.

III. EXISTING PROBLEM

India has a wide range of laws to protect children and women and their protection is increasingly accepted as a core component of social development. The challenge is in implementing the laws due to inadequate human resource capacity on the ground and quality prevention and rehabilitation services. As a result, millions of children and women are prone to violence, abuse and exploitation. Violence takes place in all settings: at home, school, childcare, institutions, work and in the community. Often violence is perpetrated by someone known to the victim. India has fairly comprehensive policy and legal framework addressing rights and protection for children, providing opportunities to ensure that all children have equal access to quality protection services. The core child protection legislation for children is enshrined in four main laws: The Juvenile Justice (Care and Protection) Act (2000, amended in 2015); The Prohibition of Child Marriage Act (2006); The Protection of Children from Sexual Offences Act (2012), and The Child Labour (Prohibition and Regulation) Act (1986, amended in 2016). Also, for women abuse, several steps are taken by the government. Some of them are:- Expansion of helplines and information sharing, Funding shelters and other safe accommodation options for survivors, Expansion of access to services for survivors, Limiting risk factors associated with violence, Modifications to family law and justice systems. Information is being shared widely through guides, resources, and advocacy targeting friends and family members. Parenting for Lifelong Health has compiled evidence-supported guidance for safe parenting during quarantine. Helplines online support platforms are being expanded or established. Italy, one of the countries hardest hit by the pandemic, is preventing “an emergency within an emergency” by advertising the 1522 helpline for violence and stalking.

Over the past five years, notable efforts have been made to set up fast track courts and deal with cybercrime against children and women. In 2019, the Protection of Children from Sexual Offences Bill was amended, stipulating stricter punishment for sexual crimes against children. Exact data on violence, abuse and exploitation is not sufficient, but overall, the nation is becoming increasingly aware of violence against children, especially sexual abuse. Several cases that may have earlier gone unnoticed, are now. Even after taking so many steps, many people are afraid of reporting and complaining to the concerned authorities. No action can be taken until a abuse is reported. Above two abuses are the ones which are paid most attention. But there are many other abuses. Consider animal abuse. Animals cannot report by themselves any harm made to them.

IV. PROPOSED SOLUTION

In all such cases, to help the victims, many socially active people take an initiative by reporting these things on popular social platforms with the use of hashtags. The



solution proposed here uses these social platforms to get abuse related posts. It is follows: - firstly, user will be provided with a website having interactive and a responsive UI. There, user will have to enter the keyword about which he wants to get the information. Here, keywords can be – child abuse, women abuse, etc. Thereafter, Web Scraping will be performed on the Twitter and the user will get the information on the same webpage in a tabular format. The tables will consist of columns like username, time of tweet, tweet description, etc. This application may also help in extracting all such abuse related posts from more than one social platform. The aim of this method is providing such information from different websites in the same place so that the authorities can get informed sooner and then they can take further actions.

V. IMPLIMENTATION

In our project we take the input from the user in the form of hashtags/keywords which will be used for the data extraction from twitter with the help of selenium using python. This whole project is developed with the help of Django framework where Selenium is a free (open-source) automated testing framework used to validate web applications across different browsers and platforms. The Selenium API uses the WebDriver protocol to control a web browser, like Chrome, Firefox or Safari. The browser can run

either locally or remotely. To scrape data from these browsers, selenium provides a module called WebDriver, which is useful to perform various tasks like automated testing, getting cookies, getting screenshots, and many more. In order to extract the information that you're looking to scrape, you need to locate the element's XPath. An XPath is a syntax used for finding any element on a webpage. Once the information is located that we have searched in the form of hashtags/keywords, it will iterate through number of paths that contains the required hashtags and collects all the require data and displays it on our web page.

VI. EXPERIMENTAL RESULT

The solution proposed in this paper was implemented on one of the most popular social media platforms – Twitter and the results were positive i.e. we are successfully extracting the following user information like user's twitter id, user's twitter name, time of tweet and his/her tweet. We were able to extract this useful information about the abuse related posts. Till now, web crawling was used to address various crime related activities in different domains. Our paper differs in terms of method and the social issue that was resolved through this case study as we are using selenium for extracting data from the twitter.



Keyword	User_ID	User_Name	Time	Tweet
Abuse	FP_Detectives	FrontPageDetectives	2022-05-17 T16:00:33.000Z	FrontPageDetectives @FP_Detectives · 4m A man killed his father's poodle "Fuzzy" and buried it in the yard, then laughed about it as he spoke to police. #Florida #dog #kill #sick #animal #cruelty #abuse #arrest frontpagedetectives.com Man Buried His Dad's Poodle, Then Laughed About It As He Told Police: Report A man in Florida faces an animal cruelty charge after police said he buried his father's poodle. Authorities said the suspect laughed about his crime.
	AlisasWish	Alisa's Wish Child & Youth Advocacy Centre	2022-05-17 T16:00:01.000Z	Alisa's Wish Child & Youth Advocacy Centre @AlisasWish · 4m Get Free Child Abuse Prevention Resources -- Including Activity Books 4 Children Pre-School - Grade 12. Just call 604-340-6914 to request copies. #VictimsWeek #ThePowerOfCollaboration #mapleridge #mapleridgebc #pittmeadowsbc #pittmeadows #abuse #child
	Hope4Siblings	Nancy Kilgore, M.S.	2022-05-17 T15:57:04.000Z	Nancy Kilgore, M.S. @Hope4Siblings · 7m #abuse #violence Many survivors of SIBLING ABUSE are still emotionally abused by their abusive sibling in adulthood. Abusive siblings are often in the upper power structures of families. GIF
Women Abuse	AustraliaTamizh	Australia Tamizhan	2022-05-15 T06:09:14.000Z	Australia Tamizhan @AustraliaTamizh · May 15 #கற்பழிப்போம் #தமிழ்ப்பதிவு #Australia Tamizhan @AustraliaTamizh · May 15 #கற்பழிப்போம் #தமிழ்ப்பதிவு #TamilBlogger #WomenAbuse 1
	Jeffrey78411848	UKPatriot	2022-05-14 T22:17:28.000Z	UKPatriot @Jeffrey78411848 · May 15 Replying to @KafirCentral I was in Sharm Egypt a few years back! 1 day per week they let the locals use the pool, the guys eying up all the western chicks in bikinis while their wife's sat there covered head to toe in total silence while the numpty men eyed up the girls! Never been back since
	SALAMMASS1	Salam Mass INC	2022-05-14 T17:33:42.000Z	Salam Mass INC @SALAMMASS1 · May 14 #பாஜக வின் ஆட்சியில் பெண்களுக்கு பாதுகாப்பு இல்லை. UPல் நடந்தது மட்டும் தான் இதில் இருக்கிறது. எண்ணிக்கை விகிதம் அதிகம். பாரதிய ஜனதாவிற்கு பதில் #பாலியல்_ஜனதா_கட்சி என மாற்றிவிடலாம். #Salam Mass INC @SALAMMASS1 · May 14 #பாஜக வின் ஆட்சியில் பெண்களுக்கு பாதுகாப்பு இல்லை. UPல் நடந்தது மட்டும் தான் இதில் இருக்கிறது. எண்ணிக்கை விகிதம் அதிகம். பாரதிய ஜனதாவிற்கு பதில் #பாலியல்_ஜனதா_கட்சி என மாற்றிவிடலாம். #womenabuse #rape #IndiaAgainstHate #BJPFailsIndia
Child Abuse	renbaz	Howard N. Zabner	2022-05-18 T12:26:57.000Z	Howard N. Zabner @renbaz · 11m Replying to @MysterySolvent #ChildAbuse This #demented and sick. #Grooming in the worst possible way. The parents should be
	deaktivat308	SaveThe Children	2022-05-18 T12:24:08.000Z	SaveThe Children @deaktivat308 · 14m #childabuse Quote Tweet Billboard Chris @BillboardChris · 13h It's not 'top surgery.' It's a radical elective double mastectomy performed on healthy girls who have been sucked into a cult by groomer schools and online influencers. https://pic.twitter.com/q4JKyrVUS2 Show
	greenshame	greenshame	2022-05-18 T12:23:20.000Z	greenshame @greenshame · 15m Some priests abused so many children they lost track - Diarmuid Martin #catholicchurch #childabuse irishtimes.com Some priests abused so many children they lost track - Diarmuid Martin Catholic Church has to be made accountable, says Mary McAleese



VII. CONCLUSION

This paper demonstrated the web crawling of twitter for various types of abuse related posts. The idea was to extract the data from twitter which is related to abuse and display it on a single webpage.

The distinctive hashtags are searched while extracting the data from twitter. This paper is envisioned to address serious social issue that are usually unreported. It describes successful execution on Twitter and can further be used for several different popular social media platforms with legal permissions, thereby, helping many victims.

VIII. REFERENCES

1. <https://kevincurran.org/dissertations/2018%20Thesis%20Evan%20Gallagher%20%20Scraping%20Websites%20for%20Law.pdf>
2. <https://blogs.unicef.org/evidence-for-action/five-ways-governments-are-responding-to-violence-against-women-and-children-during-covid-19/>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7364393/>
4. <https://www.upgrad.com/blog/web-scraping-projects-topics-for-beginners/>
5. <https://www.unicef.org/india/what-we-do/child-protection>.
6. https://www.cl.cam.ac.uk/~bjc63/tight_scrape.pdf