International Journal of Statistics and Analysis.
ISSN 2248-9959 Volume 11, Number 1 (2021), pp. 5-13

http://www.ripublication.com

Analysis of the Factors Influencing Cybercrime using Linear Regression and Correlation Analysis

Deveshi Mehta

The Shri Ram School, V-37, Moulsari Ave, DLF Phase 3, Sector 24, Gurugram, Haryana 122002, India

Abstract

Cybercrime is crime related to information technology. It has grown rapidly with the proliferation of the internet, growth in e-commerce and rise in the use of social media platforms. This paper aims to give a better understanding of the types of cybercrime and uses linear regression and correlation to analyse how these factors have influenced cybercrimes over the years in India. Secondary data was used for the analysis and a very strong correlation between each factor and the number of crimes was found. Various events which could have triggered the steep rise of cybercrime cases over the years have been discussed profusely in the paper. COVID-19's impact on the same has also been analysed in the paper.

Keywords. Cybercrime, Linear Regression, Correlation, E-commerce, Social media

1. INTRODUCTION

Crime is an act of offence that is strongly disapproved by society. Every society defines crime keeping in mind the society's values, constitution and legal framework. India has seen an alarming rise in crimes like murder, burglary, rapes, domestic violence and failure to pay taxes in the past [1]. One of the most prominent crimes in recent times is crime relating to information technology or cybercrime. The term 'cybercrime' comes from the rapid growth of the information highway that has made way for new forms of crime online. Cybercrime is a reality that India is ill-equipped to tackle. It is an unpredictable, proliferating phenomenon, which is often untraceable and has unlimited reach. A hacker from one corner of the globe can break into a system at the other end, thereby creating problems of jurisdiction, at the very least. The low costs involved and the relative ease of conducting a cybercrime have translated into an explosion of hacks

in recent years. Moreover, the attacks are growing in scope and geographical reach, with malware such as WannaCry and Petya affecting not just India but most countries across the world. India is among the hardest hit countries under this new paradigm [2].

In the post COVID-19 period, the practice of social distancing affected people in several ways: social, economic, and psychological. It altered the way people interacted with one another bringing greater focus on a virtual platform. Internet has become an important part of our lives. As India is slowly digitising, a combination of relative digital immaturity is making India susceptible to a diverse barrage of cyberattacks. India ranks fairly high on the global list in overall cybercrime levels and accounts for 3% of the systems across the world that have been infiltrated by criminals, according to global professional services firm Ernst and Young [2]. India has witnessed more than 457% rise in cybercrime incidents under the Information Technology (IT) Act, 2000 from the year 2011 to 2020. According to a recent joint study by ASSOCHAM-NEC Online shopping and wide use of "social media" are root causes of cybercrimes [3].

Increased internet penetration, growing number of mobile internet users, increased e-commerce transactions, online banking and increasing number of social media platforms are making it easy to commit cybercrime. India ranked among top five countries to be affected by cybercrime. The number of internet users grew at a CAGR of 44%, of which India is placed third after US and China [3]. During 2020, due to the pandemic and enforced lockdown, online shopping was the only sector that flourished during the lockdown. E-commerce provided customers with a convenient-contact-free shopping experience. The convenience factor led to a big impetus in online shopping. E-stores such as Amazon, eBay, Flipkart, Alibaba have become very popular. A report by Frost & Sullivan and Microsoft in 2017 confirmed that Amazon India had to face a security breach where almost 400,000 seller's data were compromised. Cybercriminals stole data records of 17 million users from Zomato, the online food delivery and restaurant aggregator in India [4].

This research paper focuses on understanding the nature and types of cybercrime and analysing the key variables driving it. The paper analyses the increase in cybercrimes due to growth of the e-commerce sector and exponential increase in the number of social network users. These factors have been analysed using the linear regression model and correlation analysis. The degree of correlation of these factors with cybercrime and regression analysis to predict cybercrime in future has also been discussed.

2. THEORETICAL BACKGROUND

Cybercrime is an act of crime that involves an internet network and a computer [5]. Most of the available literature on computer crimes focuses on computer-related fraud. Cybercrimes can be broadly categorized in two ways:

a) *The computer as a target*: using a computer to attack other computers; for example, hacking, virus attacks, and denial of service attacks.

b) *The computer as a weapon*: committing real world crime using a computer; for example, cyber terrorism, credit card fraud, and pornography.

When a common man talks of computer crime, they mostly think of those who break into computers to steal or destroy information. We can get a larger definition of cybercrime by examining the investigations of law enforcement agencies. The National Computer Crime Squad (NCCS), which concerns itself with all crimes involving computers in two or more states, considers the following to be important computer crimes [6]:

a) *Financial Crimes*: Financial cybercrimes include credit card crimes, cheating, money laundering, and online banking frauds. These have increased with increasing online banking, e-commerce and mobile usage.

b) *Cyber Terrorism*: Cyber terrorism includes denial of service attacks, hate websites and emails, and attack on service networks.

c) *Software Piracy*: Theft of genuine software through illegal copying, counterfeiting and distribution of fake products intending to pass off as original.

d) *Virus attacks*: Deliberate attacks on networks, software, email etc. using virus, worms, web jacking etc. are common forms of cybercrime. A virus is a program that modifies other computer programs by attaching itself to a file and spreads from one device to another. A worm on the other hand is a standalone program that replicates itself and makes its way throughout the network system.

e) *Drug trafficking:* Drug traffickers use latest technologies for encrypting mails and sell narcotics in the absence of personal communication between the dealer and the buyer these exchanges are more comfortable using couriers.

f) *Cyber Pornography*: This includes setting up pornographic websites and using computers to publish and material relating to pornography. Also, abusers and podophiles use the internet to reach directly or via chat rooms to sexually abuse children worldwide.

g) *Cyberstalking*: This involves following and pursuing a person or organization's whereabouts on the Internet. They could include sending messages on the victim's bulletin boards through e-mails or social networking sites. Invading someone's privacy is a form of harassment that can leave the victim scared and threatened.

h) *Online Gambling*: Many websites that offer online gambling have their servers hosted abroad. These websites are very important sites for money launderers.

i) *E-mail Spoofing and Phishing Scams*: Cyber criminals often spoof e-mails of individuals. E-mail spoofing means sending an e-mail from a particular source while it appears to have been sent from another e-mail. It is a very common cause of monetary damages. The act that attempts to obtain important information like details of credit cards and passwords by pretending to be a trustworthy entity in an electronic company is called phishing. Phishing e-mails are likely to contain hyper-links to the sites containing malwares.

3. METHODOLOGY

Secondary data was collected from Statista [7] for analysis in this paper. For analysing the effect of the E-commerce sector on the number of cybercrime cases, data from 2014 to 2020 was collected, as per availability. For understanding the effect of social media usage on cybercrimes, data was collected from 2010 to 2020. The analysis was made using linear regression and correlation.

3.1 Linear Regression

Linear regression is a model which takes into account the relationship of two variables, a dependant and independent variable. The dependant variable is plotted on the y axis and the independent variable is plotted on the x axis. A scatterplot is constructed and a line of best fit is modelled on the graph which is used for prediction and forecasting. Linear regression minimizes the sum of the squared errors to fit a straight line to a set of data points [8]. It is a statistical relationship, and association between the variables does not ascertain that they are deterministic in nature.

The dependent variable is the variable whose values we want to explain or forecast. Its value depends on something else. It is denoted by Y. The independent variable is the variable that explains the other one. Its values are independent and is denoted by X. An equation is deduced from the graph and the statistics are used to forecast new observations:

$$Y = aX + b$$

Where ‘a’ is the slope of the line of best fit and ‘b’ is the intercept. We try to have the line as close as possible to all points, and an equal number of points above and below the line. A coefficient of determination R^2 is the variation in the number of observed data points that lie on the line of best fit.

3.2 Correlation analysis

A scatterplot can determine the kind of relationship that the variables share. It can determine whether the variables are positively, negatively or not correlated at all. As shown in Figure 1(a), for a positive correlation the slope of a graph is an upward sloping one and the range of correlation is anywhere between 0.2 to +1. For a negative correlation (Figure 1(b)), the slope of a graph is a downward sloping one and the range of correlation is anywhere between -1 to -0.2. A graph with very weak to no correlation ranges from -0.2 to 0.2 (shown in Figure 1(c)). There are two main mathematicians who have proposed theories for correlation. They are Karl Pearson and Charles Spearman who have devised a method and formula for finding a degree of correlation between two variables.

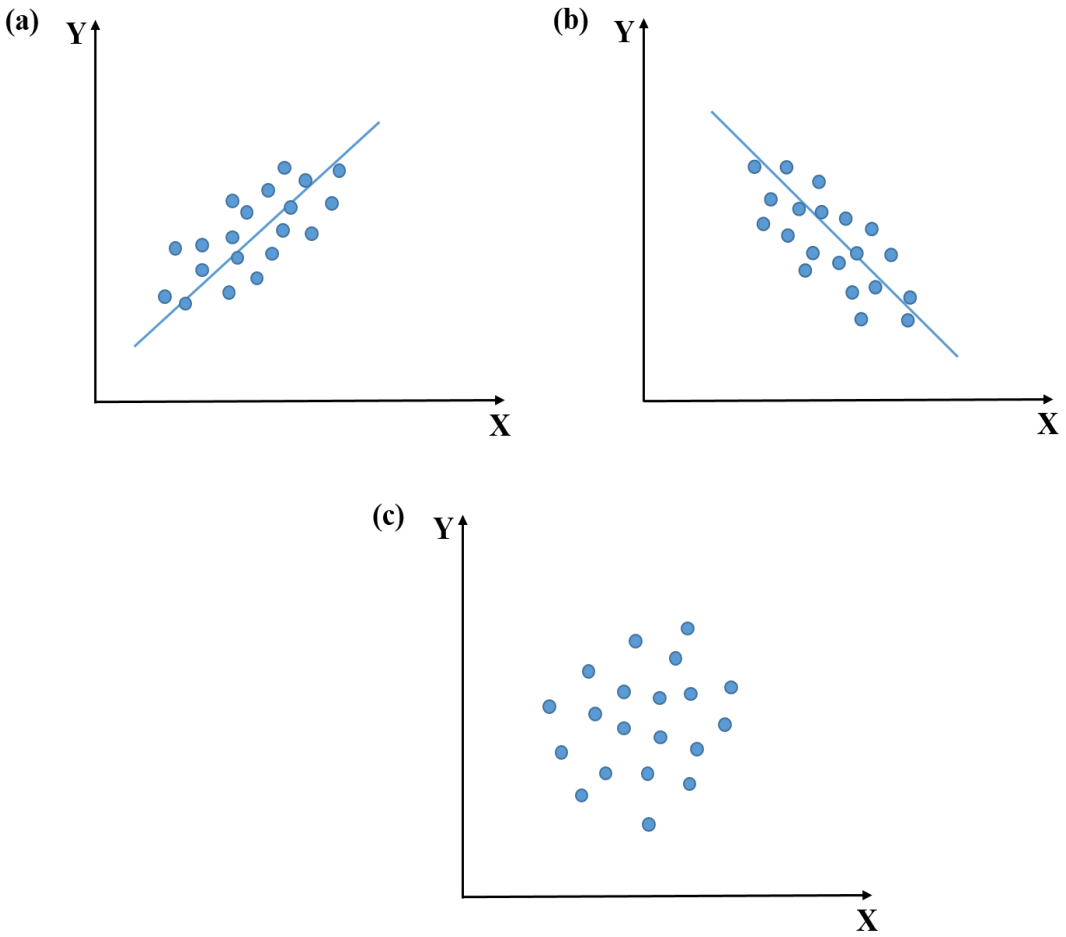

Figure 1. Schematic representation showing the types of correlation between the dependent and independent variables. (a) Positive linear correlation (b) Negative linear correlation (c) No correlation

4. RESULTS AND DISCUSSION

4.1 Effect of E-commerce on the number of cybercrimes

E-commerce functions differently from the traditional version of the trade business – with no physical presence – which is why the chances of fraud are a lot higher. Most businesses today are attracted to the convenient and cost-efficient business model an online platform provides, however, this has a flip side. Since e-retailing in India is in its initial stage, buyers and sellers are prone to getting fooled easily which has caused an increase in online criminal activity. Buyers are often fooled by fake websites due to lack of verification and are cheated on large amounts. Customers take advantage and exploit the benefits of the online system. As digital transactions are experiencing a steady growth, cyber-security has also become a concern recently.

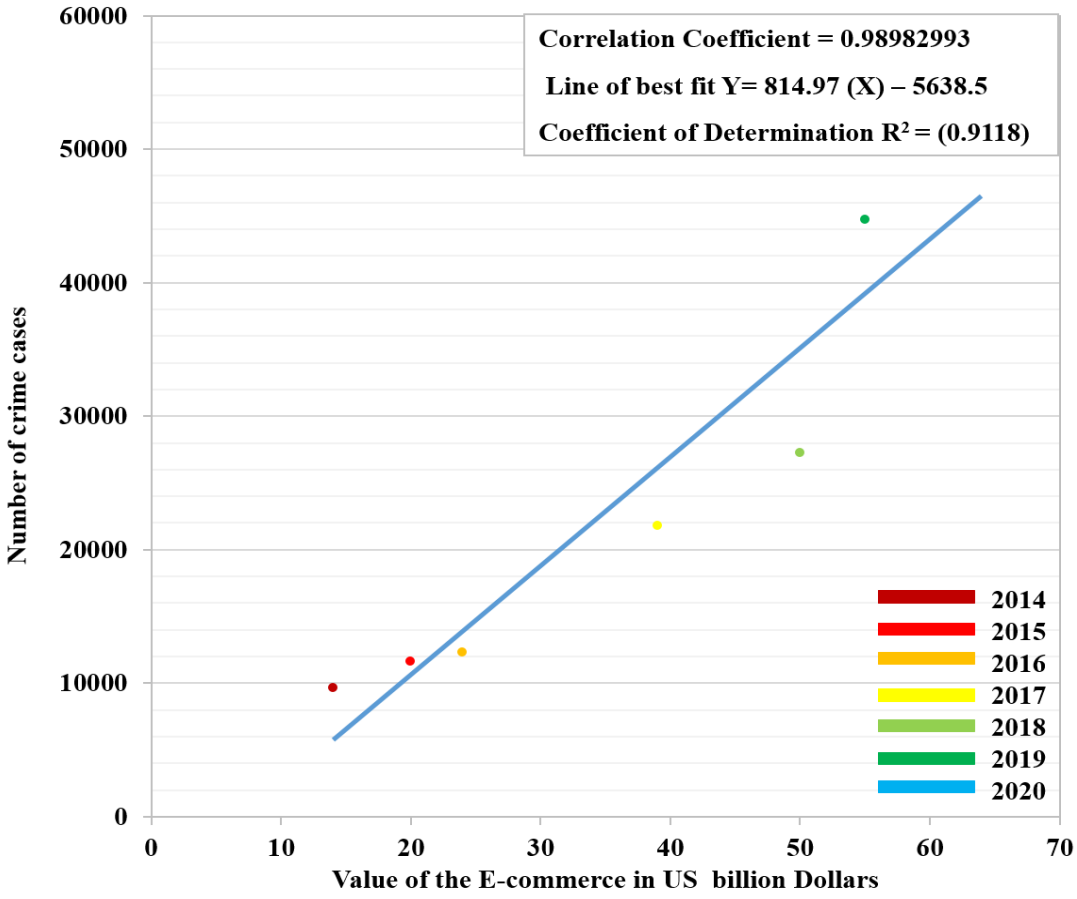

Figure 2. The relation between the number of crime cases and the value of the e-commerce

The graph in Figure 2 depicts the linear relationship between growth of the E-commerce sector in monetary terms (USD) and the number of cybercrime cases. The coefficient of correlation, found to be 0.9898 (correct to four decimal places) suggests a very strong relationship between the two. The coefficient of determination (R^2) is also high, which proves that the line of best fit has little variance between the observed values and the fitted values. There is a sharp rise in the year 2017, both in crime rates as well as in the E-commerce sector. Over 50% of cybercrimes (12,213) were done for monetary gains through fraudulent means [9]. Another important thing to note is that 2017 was the year when demonetisation was implemented. This gave the E- commerce industry a boost creating a proportionate increase in the number of cybercrime cases.

With the increase in online users, cybercrimes rose from 53,117 in the 2017 to 208,456 in 2018 [4]. In 2019, the cyber-crimes losses in the country were estimated at INR 1.25 lakh crores. The threat of cybercrime is now forcing e-commerce giants like Amazon, Flipkart and Snapdeal to trace loopholes in their existing systems. Amid the Covid-19 pandemic in 2020, the lines between the physical and digital worlds blurred, making

contactless payments and digital transactions the new norm. The Indian E-commerce market is growing steadily and is expected to be valued at $200 billion by the year 2026 [10].

4.2 Effect of Social Media on the Number of Cybercrimes

In 2019, India was the second largest country with the number of Facebook users, thirty million twitter users and over one million WhatsApp users [11]. It is expected that in the next one year, mobile internet users will be three hundred twenty million and total internet users shall be five hundred million [12]. Social media has emerged as a powerful platform that has revolutionised business transactions, commerce and the way we keep in touch with our friends. The cybercrime portal of the Ministry of Home affairs, India describes all the types of cybercrimes along with the punishments for each [13]. In 2021, the Mumbai police brought five cyber police stations under its crime branch, which a great effort to increase cybersecurity in the country [14].

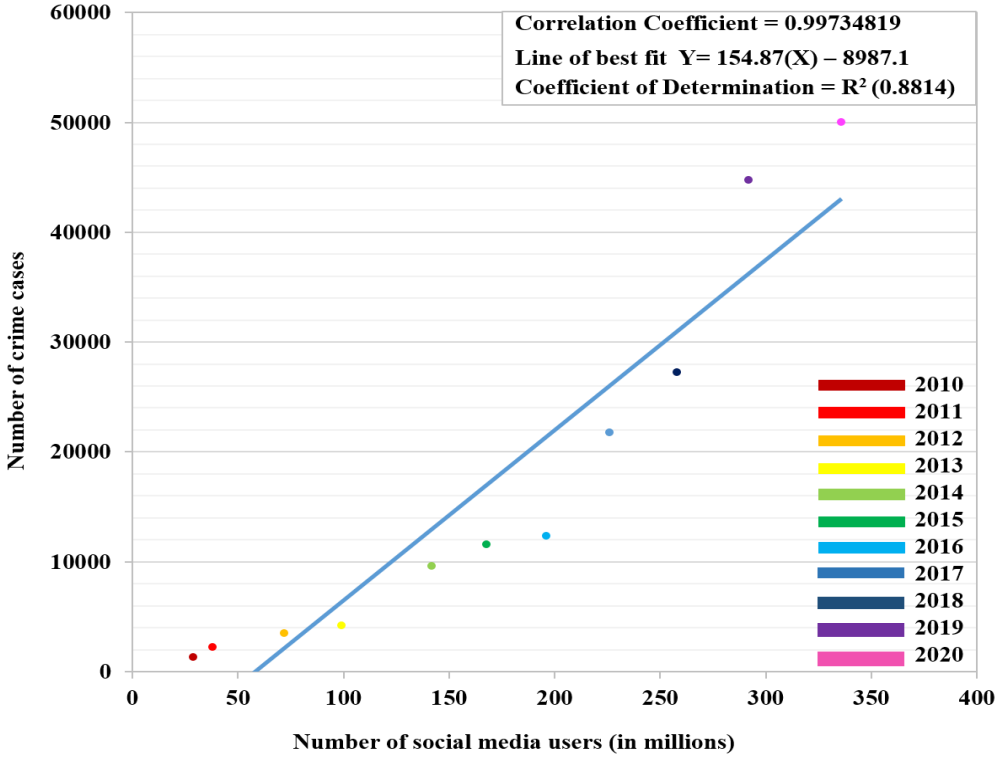

Figure 3. Graph showing the relation between cybercrime and number of social media users over a period of 10 years from 2010 to 2020.

Figure 3 depicts the relationship of the number of social media users with the number of cybercrimes. The correlation coefficient, found to be 0.9973 (correct to four decimal places), shows a very strong relationship between the two variables and proves that they are very strongly correlated. Though correlation does not imply causation, a lot of studies have shown that there has been an increase in cyber sexual harassment cases via

all the social media platforms due to increased activity by users [15]. There is a steep increase between 2018 and 2020 as the development of the various social media apps became a source of income for many. Vlogging, paid promotions and various other forms of advertisement have found their presence on apps like Instagram and Facebook. The launch of Reliance JIO and its free data services have had significant, although indirect, impact on cybercrime cases. Increased availability of internet connections and access in recent years, propelled by the central government's Digital India initiative, was directly proportional to the growth of social media users [12]. Social media platforms have become a rich platform for all the hackers, as well. Social media platforms' ceaseless focus on the acquisition of personal data has turned them into data banks that are highly attractive to cybercriminals.

The coefficient of correlation indicates that social media has a stronger correlation with cybercrime than e-commerce does, however, this is by a very small margin. As a scatterplot gives an immediate visual impression of the possible relationship between the variables, the coefficient of correlation is the numerical measurement used as a quantitative value of strength of a linear relationship. The line of best fit that has been deduced for the relationships in both the cases can be used to predict the number of cybercrime cases in the future as the other two factors vary in the coming years.

CONCLUSION

This paper analyses the growing cybercrime rates in India and how the e-commerce and social media sectors have deeply aggravated their numbers. Both the factors show high correlation values over a decades' time, from 2010 to 2020. The paper discusses various events over the years which could have influenced cybercrimes in the country. The positives and negatives of all the related events have been given importance. It is crucial for India to take cyber-security threats more seriously by strengthening cyber-security standards and implementing better security infrastructure. There is a greater need to identify and fix broken security issues as the Indian e-commerce market and social media users are expected to grow even more in the coming decades.

REFERENCES

[1] "India: Promoting internet safety amongst 'netizens.'" https://www.unodc.org/southasia/frontpage/2012/May/india_-addressing-the-rise-of-cybercrime-amongst-children.html (accessed Oct. 22, 2021).

[2] "As India digitises, cyber crime is becoming an increasingly tangible threat." https://www.consultancy.in/news/1081/as-india-digitises-cyber-crime-is-becoming-an-increasingly-tangible-threat (accessed Oct. 22, 2021).

[3] "cybercrime: India saw 457% rise in cybercrime in five years: Study, Telecom News, ET Telecom." https://telecom.economictimes.indiatimes.com/news/india-saw-457-rise-in-cybercrime-in-five-years-study/67455224 (accessed Oct. 22, 2021).

[4] "Annual Report | IBEF." https://www.ibef.org/annual-report.aspx (accessed Oct. 22, 2021).

[5] E. Ramdinmawii, S. Ghisingh, and U. M. Sharma, "A Study on the Cyber-Crime and Cyber Criminals: A Global Problem," *Int. J. Web Technol.*, vol. 004, no. 001, pp. 7–11, 2015, doi: 10.20894/ijwt.104.004.001.003.

[6] S. Kumar, "Cyber Crimes in India : Trends and," vol. 6, no. 1, pp. 25–37, 2019.

[7] "• Statista - The Statistics Portal for Market Data, Market Research and Market Studies." https://www.statista.com/ (accessed Oct. 22, 2021).

[8] M. A. Awal, J. Rabbi, S. I. Hossain, and M. M. A. Hashem, "Using linear regression to forecast future trends in crime of Bangladesh," *2016 5th Int. Conf. Informatics, Electron. Vision, ICIEV 2016*, pp. 333–338, 2016, doi: 10.1109/ICIEV.2016.7760021.

[9] N. LEENA, "Cyber Crime Effecting E-commerce Technology," *Orient. J. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 209–212, 2011, [Online]. Available: http://www.computerscijournal.org/dnload/N. Leena/OJCSV04I01P209-212.pdf.

[10] "NCRB 2017 data: Cyber crimes reached a new high in 2017 - The Hindu." https://www.thehindu.com/data/cyber-crime-cases-in-india-jumped-77-in-2017-compared-to-2016/article29889061.ece (accessed Oct. 22, 2021).

[11] "Indians 2nd largest Facebook users - BusinessToday." https://www.businesstoday.in/magazine/technology/story/facebook-india-second-largest-number-users-world-31686-2012-04-26 (accessed Oct. 22, 2021).

[12] "• Social media users in India | Statista." https://www.statista.com/statistics/278407/number-of-social-network-users-in-india/ (accessed Oct. 22, 2021).

[13] "Cyber Crime Portal." https://cybercrime.gov.in/Default.aspx (accessed Oct. 22, 2021).

[14] "Five cyber police stations brought under Mumbai crime branch, to probe fraud cases above Rs 10 lakh | Mumbai news." https://indianexpress.com/article/cities/mumbai/cyber-police-crime-branch-probe-fraud-cases-7430797/ (accessed Oct. 22, 2021).

[15] D. Sethi and S. Ghatak, "Mitigating Cyber Sexual Harassment: An Insight from India," *Asian Themes Soc. Sci. Res.*, vol. 1, no. 2, pp. 34–43, 2018, doi: 10.33094/journal.139.2018.12.34.43.