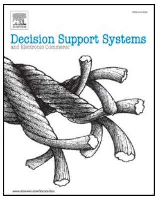
Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss



A text-mining based cyber-risk assessment and mitigation framework for critical analysis of online hacker forums

Baidyanath Biswas[a,b], Arunabha Mukhopadhyay[c], Sudip Bhattacharjee[d,*], Ajay Kumar[e], Dursun Delen[f,g]

[a] International Management Institute (IMI), Kolkata, India
[b] DCU Business School Dublin City University, Dublin, Ireland
[c] Indian Institute of Management, Lucknow, India
[d] Department of Operations and Information Management, School of Business, University of Connecticut, United States
[e] EMLYON Business School, France
[f] Center for Health Systems Innovation, Spears School of Business, Oklahoma State University, United States
[g] School of Business, Ibn Haldun University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Information security
Cyber risks
Hacker forum
Machine learning
Sentiment analysis

ABSTRACT

Online hacker communities are meeting spots for aspiring and seasoned cybercriminals where they engage in technical discussions, share exploits and relevant hacking tools to be used in launching cyber-attacks on business organizations. Sometimes, the affected organizations can detect these attacks in advance, with the help of cyber-threat intelligence derived from the explicit and implicit features of hacker communication in these forums. Herein, we proposed a novel text-mining based cyber-risk assessment and mitigation framework, which performs the following critical tasks. (i) *Cyber-risk Assessment* - to identify hacker expertise (i.e., *newbie*, *beginner*, *intermediate*, and *advanced*) using explicit and implicit features applying various classification algorithms. Among these features, *cybersecurity keywords, sharing of attachments,* and *sentiments* emerged as significant. Further, we found that expert hackers demonstrate leadership in the online forums that eventually serve as *communities of practice*. Consequently, novice hackers gradually develop their cyber-attack skills through prolonged observations, interactions, and external influences in this social learning process. (ii) *Cyber-risk mitigation* – computes financial impact for every {*hacker expertise, attack-type*} combination, and then by ranking them on a {*likelihood, impact*} decision-matrix to prioritize mitigation strategies in affected organizations. Through these novel recommendations, our framework can guide managers to decide on appropriate cybersecurity controls using an {*expected loss, probability, attack-type, hacker expertise*} metric against financial losses due to cyber-attacks.

1. Introduction

Cybercriminals have adversely impacted the global economy to billions of dollars of losses across various organizations in recent years. For instance, in December 2020, attackers conducted a large-scale breach across the users of Orion, a network monitoring product by SolarWinds. Organizations affected in the attack included top U.S. federal agencies such as the Department of Justice, U.S. Treasury, Homeland Security, and Fortune 500 companies such as Microsoft, Intel, Cisco and their own clients, as well as cybersecurity firm FireEye and many more.[1] Threat actors used a malware code named *Sunburst* and surreptitiously introduced it into the organizational networks as early as September 2019 but went undetected for over a year.[2] In another incident during January 2021, state-sponsored threat groups actively exploited four zero-day vulnerabilities in the Microsoft Exchange server, and deployed backdoors to launch widespread attacks. Some of the most targeted industries in this attack were *government and military* (23%), followed by *manufacturing* (15%), and *banking and financial services* (14%).[3] These

* Corresponding author.
E-mail address: sudip.bhattacharjee@uconn.edu (S. Bhattacharjee).

[1] Partial lists of organizations infected with Sunburst malware:https://www.zdnet.com/article/partial-lists-of-organizations-infected-with-sunburst-malware-released-online/

[2] New Findings From Our Investigation of SUNBURST:https://orangematter.solarwinds.com/2021/01/11/new-findings-from-our-investigation-of-sunburst/

[3] Exploits on Organizations Worldwide:https://blog.checkpoint.com/2021/03/11/exploits-on-organizations-worldwide/

https://doi.org/10.1016/j.dss.2021.113651
Received 15 January 2021; Received in revised form 11 June 2021; Accepted 15 July 2021
Available online 22 July 2021

incidents suggest an ever-growing trend where the likelihood of cyber-attacks are increasing in recent years, and are continuing to have a significant negative economic impact on organizations.[4] According to a recent World Economic Forum Report,[5] organizations need to use "active defence" to survive in the age of advanced cyber-threats. Therefore, cyber-attacks require proactive intervention from governmental, non-governmental, and business organizations alike.

Globally, hacker communities, also known as "dark forums", have become one-stop sites for cyber-criminals who exchange malicious technical knowledge, hacking tools and exploits before conducting cyber-attacks. These online "dark forums" are novel and promising sources of cyber-threat intelligence that firms and technology professionals can proactively scan to avert future cyber-attacks [6,46]. These forums originally belong to the "Dark Web", which can be of four primary types: *dark forums*, *internet chat boards*, *darknet markets*, and *carding shops*. This study focuses on *dark forums* as *virtual communities-of-practice* where groups of people share a mutual concern - i.e., the pursuit of malicious technical knowledge [36,46]. Armstrong and Hagel [2] have defined *virtual communities* as computer-mediated platforms where they highlight member-generated content, leading to its mutually cognitive integration of the content. To elaborate, we present the message exchange mechanism for the *Hackhound* forum[6] (Table 1), where hackers seek to expand their knowledge through continual interaction [20]. A *beginner* hacker is interested in acquiring knowledge: e.g., *David87965* participated in *Books from offensive* sub-forum to gather technical information and *Members' security* to inquire about a possible breach.

In contrast, senior hackers are more interested in sharing knowledge. For instance, *Hacker4Life* shared a malicious worm in *Black Worm Generator* sub-forum. The *Community of Practice Theory* [51] supports these behavioural traits through social learning, where an individual's knowledge acquisition is dependent mainly on peers and mutual interactions among them [13]. Such knowledge-sharing behaviour can serve as *explicit* predictors of hacker expertise, where *beginners* posted messages at a rate of $520/159 = 3.271$ per hacker versus $665/374 = 1.778$ per *newbie*.

Again, analysts can examine message-exchange mechanisms in these dark forums to reveal *implicit* predictors of hacker expertise, such as the number of cyber-security keywords published by each hacker. For instance, "intermediate" hackers, *Hacker4Life* wrote "dark worm," and *BlackArray* posted "improved security scanner," which were technically more enriching than what "beginner" hacker *googlefloober* and *David87965* wrote. Therefore, cyber-security analysts can detect hacker expertise using such meaningful combinations of *explicit* and *implicit* message-exchange features. Subsequently, firms can design mitigation strategies based on the hacker's level of knowledge (i.e. expertise) and type of attack he/she can inflict, thereby preventing future attacks. Such proactive IT risk management techniques are known as *cyber-threat intelligence* [5,6,34,45]. For instance, sensitive financial[7] and personal information[8] leaked from consumers is often available on darknet forums for sale. Credit-monitoring firms can proactively investigate these forums, extract similar information and possibly prevent large-scale data breaches in future. With this in mind, scholars and practitioners admit that a deeper understanding based on hackers' *explicit* and *implicit* message-exchange behaviour is required to determine their expertise. These are eventually needed to minimize the efficacy and extent of similar attacks in the future. Therefore, building from these current gaps and objectives of organization-level cyber-threat intelligence, we pose three research questions that are highly relevant for firms and cyber-security researchers:

- **RQ1**: What are the determinants (both explicit and implicit) of the expertise of hackers {such as *newbie*, *beginner*, *intermediate*, or *advanced*} in dark forums?
- **RQ2**: What is the likelihood of getting attacked by a hacker even after a successful cyber-threat intelligence analysis?
- **RQ3**: What will be a firm's cyber-risk mitigation strategy {expected loss, probability} when faced with different types of attacks from these hackers for {attack-type, hacker expertise}?

We seek answers to these questions by proposing a two-stage framework, as shown in Fig. 1. In the first stage, we built a cyber-risk assessment module, using *hacker-expertise* as an input that we measured by (i) quantitative features, or *explicit* (by examining participation behaviour of the hackers within the dark forum), and (ii) qualitative features, which were *implicit* (by analyzing the content of communication made by hackers). This module provides us with the probability of correctly classifying hackers into various expertise levels: *novice*, *beginner*, *intermediate*, and *advanced*. In the second stage, we built a cyber-risk mitigation module, where (i) we apply this probability to compute the expected losses arising from major attack-types which these hackers could launch if they went undetected, (ii) built a *risk-impact matrix* using the {expected loss, probability, attack-type, hacker expertise} tuple, and (iii) proposed cyber-risk mitigation strategies using the risk-impact matrix. This study found that firms are most vulnerable to phishing attacks that compromise personal and financial information [3,15], followed by virus attacks launched by midway groups of hackers such as *intermediate* and *beginners*. Based on these findings, this study proposed actionable risk mitigation strategies.

The remainder of this paper is organized as follows. Section 2 presents an overview of existing studies and theoretical premises on hacker forums and the "dark side of information technology". Section 3 explores the data and describes the methodology. Section 4 presents the modelling techniques. Section 5 presents the empirical results. Finally, section 6 discusses the research findings from the results, implications of this study, and concluding remarks.

2. Literature review and theoretical background

2.1. Background work on Hacker forums and dark-side of information technology

Scholars have examined hacker forums as a part of the literature on the "dark side of information technology" [16,18]. These forums allow hackers to exchange messages, malicious codes and other technical assets [4,22,45,46]. Often, these discussions help in breaching the computer networks of organizations and cause financial losses [7,18,40]. Primarily, there are four categories of dark web platforms that researchers have examined to extract first-hand cyber-threat intelligence [6], namely: (i) *hacker forums* or *dark forums* (i.e. technical discussion boards for hackers) [4,8]; (ii) *darknet markets* (i.e. online marketplaces selling illicit goods) [46]; (iii) *internet relay chat forums* (i.e. online chatting platforms for hackers) [5]; and (iv) *online carding shops* (i.e. platforms to sell stolen personal and financial credentials) [31]. Scholars prefer *hacker forums* for faster extraction of cyber intelligence [6], while systematic challenges plague other platforms. Subsequently, we review the recent literature in IS that has examined hacker expertise within these forums.

[4] World Economic Forum: The Global Risks Report 2020: http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf

[5] Wild Wide Web, Consequences of Digital Fragmentation: https://reports.weforum.org/global-risks-report-2020/wild-wide-web/

[6] HackerWeb Forum Collection-Hackhound Forum Dataset: https://www.azsecure-data.org/other-forums.html

[7] Credit card details worth nearly $3.5 million put up for sale on hacking forum: https://www.zdnet.com/article/credit-card-details-worth-nearly-3-5-million-put-up-for-sale-on-hacking-forum/

[8] Thousands of hacked Disney+ accounts are already for sale on hacking forums: https://www.zdnet.com/article/thousands-of-hacked-disney-accounts-are-already-for-sale-on-hacking-forums/

Table 1
Messages in the Hackhound forum indicating knowledge exchange among hackers.

Sub forum	Thread	Date	Seq#	Message	Hacker	Role
Books from offensive…	books-from	07/20/2014	1	Published in……	***3rror4o5***	A
		07/21/2014	2	Thanks You are welcome my friend	***David87965***	B
Black Worm Generator	Black-worm-generator	08/14/2014	1	Hi, here I leave you a very good worm	***Hacker4Life***	I
		08/23/2014	2	thanks for sharing the dark worms, appreciate it	***googlefloober***	B
		08/23/2014	3	You're welcome!	***Hacker4Life***	I
		12/21/2014	4	I hope it is clean. Thanks for sharing	***h4ck2k***	N
Members security	members-security	02/24/2013	1	Due to the latest events, trying to hijack accounts.	***Ravage***	E
		03/01/2013	2	Is the scanner coming back?	***BlackArray***	I
		03/01/2013	3	Well I think you should move scanner to a separate hosting	***NK2***	N
		03/01/2013	4	No, they improved security of scanner	***BlackArray***	I

N = Newbie/Novice; B = Beginner; I = Intermediate; A = Advanced; E = Expert.

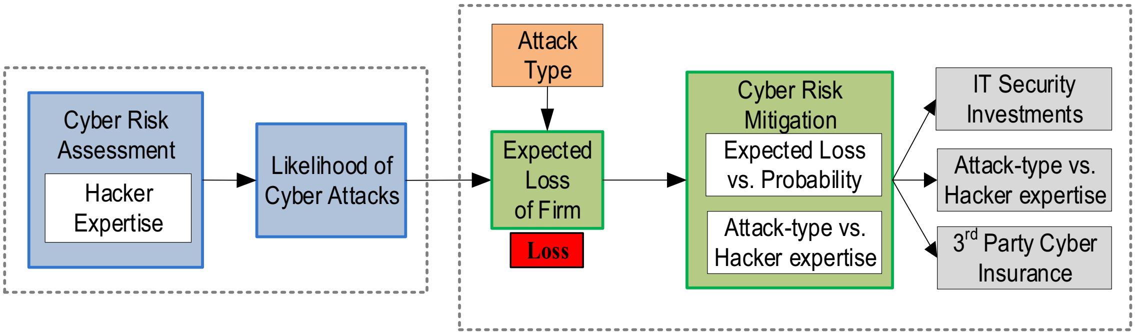

Cyber Risk Assessment
Hacker Expertise
Likelihood of Cyber Attacks
Attack Type
Expected Loss of Firm
Loss
Cyber Risk Mitigation
Expected Loss vs. Probability
Attack-type vs. Hacker expertise
IT Security Investments
Attack-type vs. Hacker expertise
3rd Party Cyber Insurance

Fig. 1. Our proposed framework with cyber risk assessment and mitigation modules.

Among the earliest studies, Benjamin and Chen [4] examined two dark forums to determine reputation score with linear regression using *average message length*, the *number of replies*, *seniority*, and *attachments*. Zhang et al. [55] extracted messages from a hacker forum, classified them into *knowledge acquisition* (e.g. questions, requests) and *knowledge provision* (e.g. answers, tutorials) and applied Support Vector Machine (SVM) algorithm to categorize these messages. Based on the above, four hacker types were identified: *guru*, *casual*, *learning*, and *novice*.

Using social network analysis, Samtani and Chen [44] identified key hackers in hacker forums for keylogging activities. They applied network-level metrics such as degree and betweenness centrality to rank top hackers in forum. Grisham et al. [22] extended prior studies [4,44] to identify key hackers in dark forums dealing with mobile malware. They employed deep learning-based text classification and social networks to study dark forums and validate their findings.

Next, Samtani et al. [45] identified emerging hacker assets (e.g. source codes, attachments) across seven hacker forums using an ensemble of machine learning and topic modelling. Additionally, they identified key hackers in these forums using network analysis similar to Samtani and Chen [44]. Subsequently, Benjamin et al. [6] proposed a DICE-E framework for data collection and evaluation from hacker forums. Based on prior literature [4,31,44], they examined four dark forums, extracted both usage (e.g. *threads*, *posts*) and message-based (e.g. *discussion of attacks*, *source codes*) features and predicted reputation scores of hackers in these forums.

Across three dark forums, Marin et al. [36] applied genetic algorithms with a set of 25 features to predict the reputation score of hackers. Their study was able to find key hackers on forums with no reputation scores or even a fragile reputation system. Finally, Huang et al. [27] proposed HackerRank, a topic-specific modified PageRank mechanism combining content analysis with social network analysis. They examined five dark forums and identified the top 50 hackers from each of them. We present a summary of these studies in Table 2.

2.2. Theoretical background

The theoretical background of this study lies primarily in the *Community of Practice Theory* [31], the *Social Exchange Theory* [10], and the *Value Co-Creation Theory* [53]. First, a *community-of-practice* describes a group of people who share a mutual concern and seek to expand their

Table 2
Summary of recent studies in IS that examine hacker expertise in online hacker forums.

Academic source	Methodology	Context/dataset	IS theory	Outcome variable(s)	Objective(s)	CRM
Benjamin and Chen [4]	OLS	Hacker Forum	CT	RS	Identify key hackers	–
Zhang et al. [55]	SVM	Hacker Forum	ST	Count of Messages	Identify key hackers	–
Samtani and Chen [44]	SNA	Keylogging Forum	GT	Network Centrality	Identify expert hackers	–
Grisham et al. [22]	SNA, DL	Mobile Malware	–	Malicious SC, centrality	Identify expert hackers, exploit	–
Samtani et al. [45]	ML, SNA, LDA	Hacker Forum	GT	Malicious SC, centrality	Exploits, key hackers, asset	–
Marin et al. [36]	ML	Hacker Forum	GT	RS	Rank hackers	–
Benjamin et al. [6]	OLS	Hacker Forum	–	RS	Identify expert hackers	–
Huang et al. [27]	CA, SNA	Hacker Forum	–	RS	Rank hackers	
This Study	ML, HLR	Hacker Forum	COP; SET	Labelled Expert Class	Expert hackers, risk mitigation	Y

ST = Statistical Learning Theory; CT = Control Theory; GT = Graph Theory; SVM = Support Vector Machine; OLS = Ordinary Least Squares Regression; SNA = Social Network Analysis; LDA = Topic Modelling with Latent Dirichlet Allocation; CA = Content Analysis; ML = Assortment of supervised machine-learning algorithms; HLR = Hierarchical Logistic Regression; SC=Source Code; DL = Deep Learning; RS = Reputation Score; CRM = Cyber-risk mitigation strategies; COP=Community of Practice Theory; SET = Social Exchange Theory.

knowledge through continual interaction [50,51]. The members of a *community-of-practice* interact informally in a connected manner and exhibit the following features. They can (a) solve problems quickly, (b) develop professional skills, (c) transfer best practices among themselves, and (d) generate an artefact or service [29,51]. Recent studies have reported that hackers regularly interact within dark forums to discuss vulnerabilities, exploits, and possible breaching mechanisms with the help of source codes, file attachments, and tutorials, which serve as artefacts [45,46]. Hence, hacker communities and their mutual interactions can fit seamlessly within a *community-of-practice* where the members exhibit social learning behaviours.

Next, *Social Exchange Theory* states that individuals seek to maximize rewards and minimize costs in any given social relationship [25] using a cost-benefit analysis. In contrast to economic exchanges, which suggest the barter of extrinsic benefits among the parties, the social exchange mechanism highlights intrinsic rewards and trust [10,19]. For instance, among online question-answering forums, users may *share knowledge* and expect to acquire it from their peers; otherwise, they aim for *peer attention* as their reward [10,25,26]. Past literature has employed *Social Exchange Theory* to examine information sharing and user behaviour in online communities [19,26,29]. Similarly, in *dark forums*, hackers seek scholarly attention among peers and therefore share key assets [45,46], use "cyber-security" relevant keywords [6], and often compose their messages using inimitable signs such as excess punctuations, URLs and font colours. These give rise to the hackers' successful dissemination of darknet knowledge and successful cyber-attacks, thereby leading to intangible rewards (i.e. sense of achievement and pride).

Value Co-creation Theory derived from the service-dominant (S-D) logic suggests that business value is *co*-created through the assimilation of resources across various members of a community, further enabled through online interactive platforms [1,48,53]. Many business firms in the automobile and consumer appliances sectors encourage the formation of communities among loyal consumers to mutually interact and produce innovative business solutions [9,48,53]. In similar lines, we observe that members of dark forums integrate resources mutually among themselves to produce community-level value that is manifested through successful artefacts (e.g. exploits, malicious codes, tutorials) and knowledge creation services (e.g. execution of cyber-attacks).

To summarize, Community of Practice Theory suggests that discussion attributes can help identify the level of seniority of a hacker in the forum, while Social Exchange Theory and Value Co-creation Theory confirm the importance of physical features and message content. Therefore, in our study, each hacker is our unit of analysis, and each of their discussion and content-based elements corresponds to a predictor by which we examine their seniority.

2.3. Identification of gaps in the current literature

The recent literature on the "dark side of IT" and hacker forums [4,6,22,31,44,55] (Table 1) had identified the presence of social networks across darknet forums with top members as central nodes. However, while identifying key actors, there has been limited methodological improvements beyond social network analysis [44,46] and simple linear regressions [4,6] to predict the reputation scores of hackers. Further, none of the recent studies has applied supervised machine-learning techniques with labelled data for classifying hackers according to their expertise.

Next, within the recent research in "dark side of IT" [5,6,22,39,40,44,45], we did not find any study that identified the presence of a multi-level ecosystem within dark forums employing these theoretical lenses (see Sub-Section 2.2). Besides, based on the theories proposed in this study, the generation of a valuable repository of malevolent knowledge by the key actors, which leads to the formation of communities of practice [50,51] through *social exchange* and *value co-creation* behaviour [1,53] is lacking in the current literature.

Considering the research on hacker forums and the dark side of IT, we identify that none of these studies has proposed suitable mitigation strategies to emphasize the cybersecurity controls and measures that organizations can employ to mitigate those cyber-risks. Further, each hacker type is unique, so are the attacks and the estimated losses arising from them. To address these gaps, our study successfully contributes to an enormously high-impact and concurrent body of knowledge on hacker forums through novel cyber-risk assessment and mitigation stages described next.

3. Data and methodology

3.1. Variables used to build the cyber-risk assessment module

We collected hacker-forum data for the *Hackhound Forum* available with AZSecure Portal, Artificial Intelligence Lab at the University of Arizona. Before pre-processing, the dataset consisted of 4242 forum posts by 834 unique hackers from October 2012 to September 2015, on a diverse set of hacking topics collected in 2015. First, we categorized the determinants of a hacker's expertise in a Darknet forum into *explicit* and *implicit* features. These were further subdivided into (i) forum-usage (*explicit*) – derived from the physical usage of the forum such as *number of threads started, number of message replies per thread*, the *sequence of messages*, and (ii) text-message content (*implicit*) – derived from the textual analysis of the online messages that were exchanged in a hacker forum such as *average positive sentiments, average negative sentiments*, and *presence of cybersecurity keywords*.

3.1.1. Determining hacker-expertise based on forum-usage features

An individual's cognitive capital consists of expertise, facility with knowledge application, and mastery of that skill, all of which increase over time as hackers interact with others via these hacker forums [50]. Therefore, the time spent on these forums ($\boldsymbol{X_1}$) can determine hacker expertise. Further, the types of discussion threads are highly determinate of hacker expertise. We also noted that expert hackers author a substantial number of messages across discussion threads ($\boldsymbol{X_2}$), similar to an online healthcare community [33,38,41].

Next, community-of-practice members enjoy extensive and frequent interactions with one another because of mutual interest [50,51]. Members can cooperate on joint exercises, exchange of ideas, and pertinent technical information. Hacker communities enjoy similar behavioural traits, and expert hackers exchange and contribute cybersecurity keywords in messages ($\boldsymbol{X_{12}}$) on these forums. They also respond to numerous questions ($\boldsymbol{X_4}$) posted by newbies and beginners in an attempt to reinforce their positions (as advanced practitioners and leaders) and reputation in the dark community [27,36]. Often, when the hacker is an initiator of a discussion ($\boldsymbol{X_5}$), or participates in the earlier sequences in a string of messages that were available for a particular topic ($\boldsymbol{X_6}$), then it is definite that the concerned malicious agent is an expert in the field of cyber-attacks [6], similar to an online healthcare community [38]. Each of these variables $\boldsymbol{X_{1,}}$ $\boldsymbol{X_{2,}}$ ….. X_6 represents the explicit message-exchange mechanisms using *forum-usage features* supported by communities-of-practice [50,51].

3.1.2. Determining hacker-expertise based on message-content features

Here, we examine the determinants of hacker's expertise based on the text messages and their content. Lengthy messages can deliver more cognitive value ostensibly and are relatively more important to the community's larger audience – be it in hacker forums or technical question-and-answer forums such as *StackExchange* [5,11,33,37]. Extant studies confirm that the message length ($\boldsymbol{X_7}$) measured by the number of words used to deliver a message strongly influences the content produced by the hacker [6,36]. Furthermore, long messages facilitate the subsequent message exchange mechanism across users in question-and-answer forums, such as Quora, where long and detailed messages receive more views and replies. Expert users and administrators, in turn, post such messages in online healthcare forums [33] [38] and knowledge communities [13]. Therefore, a similar role-based demarcation is expected in hacker forums. Further, the prevalence of sharing technical knowledge through associated artefacts such as botnets, malware, payloads, and corrupt files to poison IPs, machines, and networks using attachments ($\boldsymbol{X_8}$), is popular among online communities-of-practice.

Next, the average positive sentiment of messages ($\boldsymbol{X_9}$) and negative sentiment of messages ($\boldsymbol{X_{10}}$) can help determine the community members' expertise. Members who possess an inherent positive attitude search for helpful information and intend to provide similar feedback and answers [43]. Such behaviour is prevalent across focal members in question-and-answer forums, e.g. *StackExchange* [37], firm-level stakeholder analysis [28] and online healthcare communities [13,14]. Again, the keyword content [35] of the messages posted by a hacker is a vital determinant of their expertise. Therefore, relevant cyber-security keywords occur more commonly in an expert's message ($\boldsymbol{X_{11}}$).

Further, special characters' usage expresses emotions ($\boldsymbol{X_{12}}$) while sharing information across online forums [23,28,47]. For instance, commas, semi-colons, colons, and question marks represent the pausality feature of sentences written by a hacker, indicating the complexity of linguistics, while emoticons represent the expressivity of emotions in a sentence [6,8,47]. Therefore, we posit that emotiveness affects the hacker's expertise in a Darknet forum.

Finally, we posit that hackers intend to attain a sense of self-worth and achievement by sharing knowledge more openly and effectively with peers. Hackers are highly risk-taking individuals ($\boldsymbol{X_{13}}$), and they often seek a sense of pride and achievement ($\boldsymbol{X_{14}}$) apart from accomplishing financial gains during a cyber-attack. These behavioural cues are more robust for an expert member in a hacker forum, thereby allowing us to study their effects on the subsequent identification of hacker expertise in a Darknet forum. Each of these variables $\boldsymbol{X_7}$, $\boldsymbol{X_8}$ ….. $\boldsymbol{X_{13}}$, $\boldsymbol{X_{14}}$ represents the implicit knowledge exchange and learning mechanism using *message-content features* supported by the *Community of Practice Theory* and the *Social Exchange Theory*.

Few implicit variables, such as *average length* (in words), *presence of punctuation symbols*, *sense of achievement*, and *risk-attitude,* were derived from linguistic cues and were generated using the LIWC (Linguistic Inquiry and Word Count) software [42]. Numerous studies confirm the validity of LIWC that examine online textual data to infer psychometric properties [13,30]. Furthermore, we calculated the cognitive dimensions embedded in the hacker messages - (i) *achievement* (such as *earn*, *hero*, *win*, *victory*), and (ii) *certainty* (such as *always*, *never*, *sure*) using LIWC [42].

We performed sentiment analyses using SentiStrength [43,47] to generate the average positive sentiment and average negative sentiment of each hacker's messages. SentiStrength uses non-lexical information and rules to detect sentiment strength from short, informal English texts. We chose SentiStrength over LIWC because it offers a scale-based (−5 to +5) measurement of positive and negative sentiments from textual data, rather than a simple count or percentile generated by LIWC.

3.2. Feature-engineering and variable transformation for the risk-assessment module

We observed that some determinants exhibited varying scales and much higher ranges than others, such as X_1, X_2, X_7, and X_{13}. Further, some variables suffered from high standard deviations, such as X_9. Therefore, to improve the empirical results and ensure the accuracy of coefficient estimations from the empirical models, we normalized the research variables and log-transformed those before model fitting (see Table 3).

3.2.1. Content analysis of hacker discussions to find significant cyber-security keywords

Next, we present the overlap scoring mechanism followed by the generation of opinion scores using sentiment analysis. These results are based on a combination of the term-frequency (TF) and inverse-document-frequency (IDF) to produce a normalized composite weight (TF-IDF) for each term in a hacker corpus, from which we built the *overlap score* as the weighted average of TF-IDF-s assigned to an individual hacker for using significant cybersecurity keywords [35], as illustrated in Tables 5 and 6. For instance, the keywords *backdoor* and *bot* for hacker *x58* are assigned TF-IDF scores of 0.937 and 0.597 compared to a TF-IDF score of 0.105 for the keyword *analysis* for hacker *Ravage*. Therefore, those who speak fewer words but constitute significant keywords will rank higher than those who talk relatively unimportant keywords [35].

3.2.2. Sentiment dictionary for relevant cybersecurity keywords ($\boldsymbol{X_9}$ and $\boldsymbol{X_{10}}$)

Next, we created our own lexicon consisting of positive and negative keywords from the generated list of significant cyber keywords and assign a relative weight to each of them. Finally, we appended this set to the list of pre-defined English keywords in *SentiStrength* [47] to extend it. In this manner, we created a domain-specific dictionary for cybersecurity-related analysis, where we categorized the "cyber attack-related" words as negatively polarized and "cyber risk-mitigation" related words as positively polarized. For instance, we assigned a score of (−3) to *virus*, *malware*, and *crypter*, (−2) to *bot*, *overflow*, *backdoor*, (−1) to *hide*, *key*, *socket*; 0 to *login*, and so on. In this way, we built a novel sentiment dictionary for application in the proposed framework, shown in Fig. 2.

3.3. Variables used to build the cyber-risk mitigation module

Then, we compute the expected loss suffered by an organization during a cyber-attack launched by a hacker in case of erroneous detection. Financial loss due to cyber-attacks can arise from different types of attacks: virus (V), denial-of-service (Dos), financial fraud (FF), system penetration (SP), theft of proprietary information (TPI), and unauthorized access to information (UA) [40]. Table 7 presents the descriptive statistics for financial loss in (in $ "000") suffered due to various cyber-attacks. Mukhopadhyay et al. [40] computed the metrics using survey data collected by the Computer Security Institute–Federal Bureau of Investigation (CSI–FBI). Further, based on the "number of employees" reported by the CSI-FBI survey respondents, it is evident that the survey data represents large-scale corporations as well as small and midsized businesses. We also validated the values from the loss distribution with leading cybersecurity reports published by IBM Ponemon,[9] Verizon[10] and other cybersecurity analysis firms from industry.

4. Empirical modelling

4.1. Cyber-risk assessment: probability computation for detecting expert hackers

We applied classification algorithms to offer a baseline performance evaluation for our multi-class hacker taxonomy problem. Due to the ordinal nature of the operationalized variables, we used **M1a**: k-Nearest Neighbor (*k*−NN), **M1b**: CART (Classification and Regression Tree) [12], **M1c**: Ensemble Boosted Tree [12], **M1d**: Multinomial Logit, and **M1e**: Hierarchical Logit in MATLAB. The generalized form of detection probability (or classification accuracy) is: $p(\mathrm{Y} = \mathrm{j} | \mathrm{X}_1 = \alpha_1, \mathrm{X}_2 = \alpha_2, \ldots, \mathrm{X}_{14} = \alpha_{14})$ where X_1, X_2……. . X_{14} are the predictors (see Table 3) of the target hacker class "Y" such that j = 1, 2, 3, 4 denotes the four levels {*newbie*, *beginner*, *intermediate*, and *advanced*}.

[9] IBM Cost of a Data Breach Report 2020: https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/

[10] Verizon Data Breach Investigations Report (2021): https://www.verizon.com/business/resources/reports/dbir/

Table 3
Brief description of the variables used for the risk-assessment module in our proposed framework.

Variable	Type	Brief description	Literature source	Theory	Transform	Mean	S.D.	Max.	Min.
Independent									
Forum Usage									
X_1	E	Days spent in the forum (N)	[6,8]	COP	$\log(X_1)$	6.875	0.268	7.431	6.420
X_2	E	Number of threads participated (N)	Developed from [6]	COP	$\log(X_2)$	1.244	0.366	3.463	0.000
X_3	E	Number of messages posted (N)	[6,8]	COP	$\sqrt{(X_3)}$	1.792	1.591	22.136	1.000
X_4	E	Number of message replies per thread (N)	Developed from [36]	COP	–	1.268	0.966	8.000	0.000
X_5	E	Count of discussion threads initiated (N)	Developed from [6]	COP	–	2.132	4.647	72.000	1.000
X_6	E	Average sequence of messages in discussions (N)	Self-Developed	COP	–	5.382	45.017	20.000	1.000
Message Content									
X_7	I	Average length (in words) of messages (N)	Developed from [43]	SET	$\log(X_7)$	3.128	0.210	0.582	0.358
X_8	I	Sharing of attachments (N)	[6,8]	SET	$\log(1+t)$	0.013	0.037	0.693	0.000
X_9	I	Average positive sentiment of messages (N)	Developed from [43]	SET	–	4.125	0.213	5.000	0.000
X_{10}	I	Average negative sentiment of messages (N)	Developed from [43]	SET	–	−3.833	0.110	0.000	−5.000
X_{11}	I	Cybersecurity keywords (Overlap Score) (N)	Developed from [35]	SET	–	0.431	0.083	2.978	0.000
X_{12}	I	Punctuation symbols {@,$,!, *, +} (N)	[6,8]	SET	$\log(X_{13})$	1.427	1.187	6.001	0.000
X_{13}	I	Sense of Achievement (N)	Self-Developed	SET	–	4.573	2.737	100.000	0.000
X_{14}	I	Risk-Attitude (N)	Self-Developed	SET	–	8.761	2.186	50.000	0.000
Dependent		No. of observations		Training	Testing				
Y		Advanced (C)	50	40	10				
		Intermediate (C)	80	64	16				
		Beginner (C)	160	127	33				
		Newbie/Novice (C)	376	300	76				

N=Numeric; C=Categorical; E = Explicit; I=Implicit; COP = Community of Practice Theory; SET = Social Exchange Theory; $t = \frac{X_9}{\sum X_9}$.

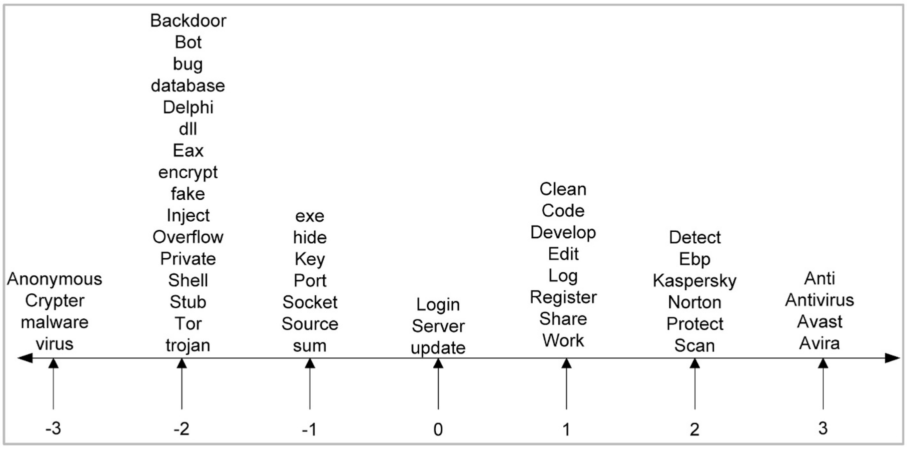


Fig. 2. Proposed sentiment dictionary for positive and negative keywords.

We applied the above classification algorithms (M1a, M1b, M1c, M1d, and M1e) to classify the target classes in the original *Hackhound* dataset - {Intelligence Service; Expert; Advanced; Advanced Member; Intermediate Member; Intermediate; Member; Beginner; and Newbie}. Based on the preliminary results from the explanatory models, we observed that the classification algorithms needed fine-tuning due to *imbalanced* target classes in the original *Hackhound* dataset. To rectify this discrepancy, we applied techniques such as (i) merging of smaller but contiguous classes (where relevant), (ii) considered precision, recall, and F-measure, instead of classification accuracy [24]. In this manner, we achieved the following revised four target classes: (i) *Intelligence Service*, *Expert*, *Advanced*, and *Advanced Member* hackers were aggregated as *Advanced (Class IV);* (ii) *Intermediate Member, Intermediate,* and *Member* hackers were aggregated as *Intermediate (Class III)*, while (iii) *Beginner (Class II)*, and (iv) *Newbie/Novice (Class I)* remain unchanged. As a result, the four classes added up to 666 unique hackers, as shown in Table 3. Next, we split this dataset following an 80:20 training-testing ratio [40] to protect the classification algorithms against overfitting and improve consistency [24].

We computed the pairwise correlations among the input variables and presented them in Table 4. Then we checked whether the variance inflation factor (VIF) stayed within the permissible limit of 10. The pairwise correlation values among the input variables ranged between 0.701 and − 0.569, while the VIF values varied from 2.145 to 1.033. Thus, the pairwise correlations and VIFs for the variables used in this study were well within the allowable ranges.

4.2. *Cyber-risk mitigation module: Risk quantification and mitigation strategies*

In the case of incorrect identification of expert hackers, the firm faces potential monetary loss. Therefore, we need to compute this expected loss [40] that may accompany significant attack types. It is given by the product of *mean of loss distribution* and the *probability of attack* [40], where the *likelihood of an attack* is given by:

$$p_a = 1 - p(\mathrm{Y} = \mathrm{j}|\mathrm{X}_1 = \alpha_1, \mathrm{X}_2 = \alpha_2, \dots, \mathrm{X}_{14} = \alpha_{14}) \quad (1)$$

$$\text{Expected Loss } (\mathrm{E_L}) = p_a {}^{*} \mu_{\mathrm{L,m}} \quad (2)$$

where $\mu_{L,\,m}$ denotes the loss suffered by a firm for attack-type m = *virus*,

Table 4
Pairwise correlations among variables used in our proposed framework.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
VIF	1.088	1.221	1.428	1.614	1.073	1.525	1.250	1.800	1.422	1.126	1.118	2.145	1.713	1.033
X_1	1.000													
X_2	0.021	1.000												
X_3	0.042	0.493*****	1.000											
X_4	0.013	0.133**	0.122**	1.000										
X_5	−0.042	0.166**	0.701**	0.028	1.000									
X_6	0.036	0.178**	0.058	0.206**	−0.024	1.000								
X_7	−0.185**	−0.125**	−0.062	−0.091*	−0.011	0.016	1.000							
X_8	0.023**	0.164**	0.054	0.209**	−0.022	0.612**	−0.023	1.000						
X_9	0.067	0.392**	0.283**	0.200**	0.096*	0.334**	−0.067	0.442**	1.000					
X_{10}	0.014	0.042	0.012	0.076	−0.019	0.164**	−0.063	0.116**	0.043	1.000				
X_{11}	0.033**	0.001	0.002	0.023	0.001	−0.034	0.022	0.003**	−0.041	0.137*	1.000			
X_{12}	−0.441**	−0.414	0.605	−0.254*	−0.112	−0.056	0.010*	−0.117*	0.024	−0.044*	0.403	1.000		
X_{13}	0.010	0.017	0.023	−0.055*	0.017**	0.012	−0.036	0.030	0.511	0.234	−0.077*	0.293**	1.000	
X_{14}	0.042**	0.016*	0.223*	0.110	0.633*	0.413	−0.017**	0.114	0.424**	−0.569	0.121	−0.135*	0.043**	1.000

[*] *$p < 0.1$.*
[**] *$p < 0.05$.*
[***] *$p < 0.01$ (2-tailed).*

Table 5
Representative hacker messages from the Hackhound forum.

Hacker	Message
h4ck2k	What content hav u uploaded??? Is it a modification of the android dev **toolkit** or has to be **analyzed**??
3rror4o5	Please autorun the **code**!! Then it will show…
NK2	this is out of my knowledge:(:(very nice work ummm, does it support .net **executables**?
x_h0rr0r_x	Hey 3rror4o5.I hav autorun the **codes**. Alright. It still is cyber software made by blackhats. Now wat to do nxt?
HttP-NuKe	PortEx is a library aimed at Java developers and reverse engineers. It enables you analyze Portable **Executable** files and has a special focus on malware analysis.
Ravage	Practical **Malware Analysis** The Hands-On Guide - EBooks - HackHound
x58	Why not add a footprint to the **bot**/panel or small **backdoor**. So you can find and kill it easily afterwards when people use it for real. I bet kids won't even verify the panel and just use it. Btw xylitol does that too.

DoS, *FF*, *SP*, *TPI*, *UA* [40]. We then rank them based on the *severity of attacks* and the *financial impact* (measured by the expected loss) for each of the attacks that each hacker group can inflict.

5. Results

5.1. Results from the cyber-risk assessment module

Table 8 compares the model-building results from the six classification algorithms: the hierarchical logit classifier (M1e) performs best at 84.852% overall accuracy, followed by the multinomial logit (M1d) 83.025% accuracy. The CART-based decision tree (M1b) achieves an overall accuracy of 72.288%, the k-nearest neighbor algorithm (M1a) at 71.453%, and the boosted tree algorithm (M1c) performs at 81.146% overall accuracy. We compared our results with prior studies and found that none of then had applied classifiers to examine labelled hacker classes (please see Literature Review Table 2). Instead, most studies predicted numerical values of reputation scores. For experiments with online hacker assessment, our performance metrics were far superior than accuracy of 64.00% with K-NN by Samtani et al. [45]; adjusted R^2 values of 57.38% with linear regression by Benjamin and Chen [4] and 52.99% by Benjamin et al. [6]; accuracy values of 75% and 79% with SVM by Zhang et al. [55]. Table 9 reports the top significant features based on the *mean decrease of the Gini Index*. Results show X_{11} (cybersecurity keywords), X_8 (sharing of attachments), and X_4 (replies per thread) are the top three significant predictors, while X_1 (days spent in the forum) and X_{12} (punctuation symbols) are the least important. Later, these values corroborate with predictors' choice in our hierarchical logit

Table 7
Central tendencies for financial loss distributions [**40**].

	Virus	DoS	FF	SP	TPI	UA
Mean	2869	729	3870	675	5869	1460
Std. Dev.	2869	729	3870	675	5869	1460

Table 8
– Comparison of algorithms (using training data).

Classifier	Overall accuracy (%)	Metrics from prior studies (%)
k-NN (M1a)	71.453	Accuracy = 64.00 [45]
CART (M1b)	72.288	–
Boosted Tree (M1c)	81.146	–
Multinomial Logit (M1d)	83.025	–
Hierarchical Logit (M1e)	84.452	–
Linear Regression	–	Adj. R^2 = 57.38 [4]; 52.99 [6]

Table 6
Generation of overlap scores (X_{11}) from the TF-IDF matrix for each hacker.

Hacker	Keyword								
	Role	*Analyze*	*Malware*	*Toolkit*	*Code*	*Executable*	*Backdoor*	*Bot*	Overlap
x58	A	0.000	0.000	0.000	0.000	0.000	**0.937**	**0.597**	1.534
Ravage	E	**0.105**	**0.685**	0.000	0.000	0.000	0.000	0.000	0.790
HttP-NuKe	A	0.000	**0.470**	0.000	0.000	**0.060**	0.000	0.000	0.530
3rror4o5	A	0.000	0.000	0.000	**0.440**	0.000	0.000	0.000	0.440
NK2	E	0.000	0.000	0.000	0.000	**0.321**	0.000	0.000	0.321
h4ck2k	N	**0.053**	0.000	**0.148**	0.00	0.000	0.000	0.000	0.201
x_h0rr0r_x	N	0.000	0.000	0.000	**0.150**	0.000	0.000	0.000	0.150

A = Advanced; E = Expert; N = Newbie.

Table 9
Top variables based on *Mean Decrease in Gini* for Multinomial Logit.

Variable	X_{11}	X_8	X_4	X_9	X_{10}	X_6	X_1	X_7	X_{12}
Importance Score	35.723	29.810	10.055	9.011	8.450	5.229	4.872	1.016	0.417

Table 10
Coefficient estimates of the hierarchical logit classifier using significant variables.

Var.	$Prob\left(\frac{y \in N}{y \in \{B, I, A\}}\right)$			$Prob\left(\frac{y \in B}{y \in \{I, A\}}\right)$			$Prob\left(\frac{y \in I}{y \in A}\right)$		
	Coeff.	Odds	S.E.	Coeff.	Odds	S.E.	Coeff.	Odds	S.E.
Const.	2.539***	12.664	0.003	2.612***	13.628	0.001	1.830**	6.234	0.002
X_1	−0.081	0.922	0.001	−0.008	0.992	0.023	−0.047	0.955	0.005
X_4	0.232**	1.261	0.040	0.688***	1.990	0.006	0.632***	1.882	0.015
X_6	0.026*	1.026	0.017	0.129*	1.138	0.019	0.110**	1.116	0.018
X_7	0.160*	1.174	0.051	0.151***	1.099	0.066	0.095**	1.163	0.058
X_8	0.542***	1.719	0.008	1.033***	2.809	0.015	1.046***	2.846	0.011
X_9	0.132**	1.141	0.031	0.610**	1.840	0.023	0.524**	1.688	0.027
X_{10}	0.056**	1.058	0.017	0.341**	1.406	0.004	0.284**	1.328	0.008
X_{11}	0.734***	2.083	0.002	1.650***	5.207	0.013	1.512***	4.975	0.005
X_{12}	−0.325	0.723	0.080	−0.227	0.797	0.002	−0.333	0.717	0.013

No. of Observations = 666; AIC = 222.043; Residual Deviance = 198.041.
Note: N ≡ Newbie; B ≡ Beginner; I ≡ Intermediate; A ≡ Advanced; *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 11
Performance metrics of hierarchical logit classifier (test data).

	Class	Predicted								
		N	B	I	A	Total	Pr	Rc	F1	(1-Rc)
Actual	Newbie (N)	72	2	1	1	76	0.900	0.947	0.923	0.053
	Beginner (B)	7	21	4	1	33	0.840	0.636	0.724	0.364
	Intermediate (I)	1	2	12	1	16	0.667	0.750	0.706	0.250
	Advanced (A)	0	0	1	9	10	0.750	0.900	0.818	0.100
	Total	80	25	18	12	135				

Pr = Precision; Rc = Recall; F1=F1 Score.

Table 12
Expected loss calculation (US$ "000") due to a security breach.

	Failure rate	Expected loss for each attack-type						Severity rank
		Virus	DoS	FF	SP	TPI	UA	
Newbie	0.053	1510	384	2037	355	3089	768	4
Beginner	0.364	10,433	2651	14,073	2455	21,342	5309	1
Intermediate	0.250	7173	1823	9675	1688	14,673	3650	2
Advanced	0.100	2869	730	3870	675	5869	1460	3

FF = Financial Fraud; SP=System Penetration; DoS=Denial of Service.
TPI = THEFT of Proprietary Information; V=Virus; UA = Unauthorized Access.

classifier (M1e) for subsequent analysis and testing.

5.1.1. *Hierarchical logistic classifier with significant variables (M1e)*

Table 10 presents the hierarchical logistic classifier (M1e), built with significant variables only. Hierarchical multinomial regression models are extensions of binary regression models but based on conditional binary observations. Eqs. (3), (4), and (5) describe the three models that are built from the hierarchical logistic classifier using the training dataset. First, in Eq. (3), the coefficient estimate for X_{11} is 0.734, which indicates that everything else remaining constant, per unit change in the *use of cybersecurity keywords,* can increase the likelihood of tracing a hacker of higher expertise (i.e., *beginner*, *intermediate*, or *advanced*) from a *newbie*, by exp.(0.734) = 2.083 times. Next, in Eq. (4), the coefficient estimate for X_7 is 0.151, which indicates that everything else remaining constant, per unit change in the *average length of messages,* leads to the likelihood of tracing a hacker of higher expertise (i.e., *intermediate*, or *advanced*) from a *beginner*, by exp.(0.151) = 1.099 times. Similarly, in Eq. (5), the coefficient estimate of 1.046 for X_8 indicates that everything else remaining constant, per unit change in the *sharing of attachments,* can increase the likelihood of being an *advanced* hacker than an *intermediate* by exp.(1.046) = 2.846 times.

$$ln\left(\frac{y \in N}{y \in \{B, I, A\}}\right) = 2.539 - 0.081X_1 + 0.232X_4 + 0.026X_6 + 0.160X_7 + 0.542X_8 + 0.132X_9 + 0.056X_{10} + 0.734X_{11} - 0.325X_{12} \tag{3}$$

$$ln\left(\frac{y \in B}{y \in \{I, A\}}\right) = 2.612 - 0.008X_1 + 0.688X_4 + 0.129X_6 + 0.151X_7 + 1.033X_8 + 0.610X_9 + 0.341X_{10} + 1.650X_{11} - 0.227X_{12} \tag{4}$$

$$ln\left(\frac{y \in I}{y \in A}\right) = 1.830 - 0.047X_1 + 0.632X_4 + 0.110X_6 + 0.095X_7 + 1.046X_8 + 0.524X_9 + 0.284X_{10} + 1.512X_{11} - 0.333X_{12} \quad (5)$$

5.1.2. *Performance metrics of the hierarchical logit classifier*

Table 11 reports the Accuracy, Precision, Recall, and F1 scores for each hacker type from the prediction algorithms. Our risk-assessment module identified the significant features (presented in Table 10) with the estimates of the hierarchical logit classifier. Results show that X_{11} (cybersecurity keywords), X_8 (sharing of attachments), X_4 (message replies per thread), X_9 (positive sentiments), and X_{10} (negative sentiments) are the top five significant predictors consistently across the four classes. In contrast, X_1 (duration), X_7 (average length of messages), and X_{12} (usage of punctuation symbols) are not significant at all. These values also corroborate the choice of predictors in the variable importance scheme measured by the *mean decrease of the Gini* (Table 9). The classifier performs with high AUC values of 0.93, 0.95, and 0.99 for *Intermediate*, *Newbie,* and *Advanced* hackers while operating at a moderate AUC of 0.79 for *Beginner* hackers.

The classifier M1e misclassifies one advanced hacker as intermediate with an associated probability of 0.409 for an *advanced* hacker. Further, from Fig. 3, our classifier M1e can successfully detect 12 out of 16 *intermediate* hackers, operating at a recall of 75%. However, it misclassifies two *intermediate* hackers as *beginners* with an associated probability of 0.190 and 0.248, respectively. The classifier M1e also marks the remaining two *intermediate* hackers as *newbie* and *advanced* hackers with an associated probability of 0.206 and 0.359, respectively.

5.2. *Results from the cyber-risk mitigation module*

Through this risk-mitigation exercise, CTOs can safeguard from financial losses that a particular type of cyber-attack can inflict. Based on extant literature and anecdotal evidence,[11] we find that expected loss distributions are highly skewed with long tails [40]. Table 12 shows the expected loss for each hacker category and attack type. Using these metrics, we propose a 2-by-2 risk- impact strategy-map (Fig. 4) based on the {*expected loss*, *probability*, *attack-type*, *hacker expertise*} metric to gauge different hacker types and attacks (i.e. Virus, DoS, FF, SP, TPI, and UA). We find that TPI attacks launched by beginner and intermediate hackers are most severe. The financial fraud and virus attacks by beginner and intermediate hackers follow next. In this manner, this study proposes a detailed risk-mitigation exercise, where we find that our cyber-threat intelligence is relatively effective in detecting *advanced* and *newbie* hackers, rather than the midway groups *intermediate* and *beginners.*

6. Discussion and conclusions

6.1. *Discussion of research findings from results*

We discuss the research findings based on the results of the Hierarchical Logit Classifier presented in Table 10 and Eqs. (3)–(5). Among *forum usage* features, we find that *days spent in the forum* is insignificant in determining hacker expertise across all levels of expertise ($\beta = -0.081$; $\beta = -0.008$; $\beta = -0.047$). Our findings coincide with Benjamin et al. [5,6] and Chen et al. [13], who analyzed knowledge contribution behaviour in an online knowledge exchange community. However, Samtani et al. [45] found that top-ranked hackers were typically senior members and published many forum messages. We explain the counter-intuitive results as follows. Often, an expert hacker may join a discussion board late but begins to disseminate knowledge and participates in information sharing as soon as they enter. In contrast, newbie and beginner hackers will often remain dormant to gain more experience and knowledge before commencing technical queries.

Next, we find that *the number of message replies per thread* published by a hacker is highly significant to classify expertise ($\beta = 0.232^{**}$; $\beta = 0.688^{***}$; $\beta = 0.632^{***}$). Whereas previous studies have simply examined the volume of messages posted by each hacker [5] [6], we computed the messages published by a hacker in each discussion thread. Additionally, observing the coefficient estimates, we found that the number of message replies was more effective for higher groups such as *advanced* than for *beginners*. Contributing to the previous body of literature on the "dark side of IT" and hacker forums [6,27,36,37], our study revealed that a hacker's expertise level is often dependent on the *number of message replies per thread* and discussions to which they contribute. This finding matches studies that examined question-and-answer forums such as *StackExchange* [11] [37], product reviews on e-commerce platforms [26], online healthcare communities [33,38,41], and knowledge communities such as Wiki forums [13]. Such behaviour is also reflected in various forums that discuss online product reviews, reported in recent IS research [9,30,43,52,54]. For instance, Yang et al. [54] found a direct relationship between the retail prices and valence of online reviews on an e-commerce platform. Further, in online health communities, Park et al. [41] reported the positive role of messages replied across various threads while other users read them.

Then, we examined the effect of the *sequence of messages in discussion threads* published by hackers in a dark forum ($\beta = 0.026^{*}$; $\beta = 0.129^{**}$; $\beta = 0.110^{**}$). While Benjamin et al. [6] reported the *number of threads started* as a predictor of expertise, recent literature in cybersecurity and dark forum analysis have failed to recognize the *sequence* of message publication as a crucial predictor. Among the closest to this finding, Huang et al. [27] and Marin et al. [36] found that discussions started by key hackers were relevant and enabled the identification of expert hackers. Our findings are similar to Mousavi et al. [38], who reported the presence of an "order effect" in online healthcare communities, where the first answer in a forum post by an expert positively influenced subsequent answers for a particular question.

Among *message-content features*, we find that the effect of *the average message size* posted by a hacker is a highly significant predictor of its expertise ($\beta = 0.160^{*}$; $\beta = 0.151^{***}$; $\beta = 0.095^{**}$). Our finding resonates with prior studies from cybersecurity research in IS, where features such as "length of replies" [5,27,36] and the "length difference" [27] emerged significantly. Similar results were also reported in online knowledge platforms [37], healthcare communities [14,33,38,41], stakeholder analysis [28], and online reviews in electronic marketplaces [9,30,43,52,54]. Especially, our findings are consistent with the measurement of absolute text length (given by *word count of reply*) and relative text length (given by the *ratio between interactive texts of host and guest's comments*) proposed by Wu et al. [52]. In addition, we computed message size by the average count of words used in the messages per hacker, which we calculated excluding special characters, and in contrast to Benjamin et al. [6]. This interesting finding also implies that communication among hackers in dark forums might not follow the usual semantics of the English language.

Then, we find that *the number of attachments shared in each message* is a significant predictor of expertise, particularly with the elite hackers ($\beta = 0.542^{***}$; $\beta = 1.033^{***}$; $\beta = 1.046^{***}$). Our findings are congruent with a few recent studies [5,45], who have presented that sharing key assets such as malicious source codes, video tutorials, cracked software, and exploits are the behavioural traits of an expert hacker (see Table 1). For instance, the *BlackPOS* malware, which hackers had used in Target and Home Depot breaches, was distributed in darknet forums before the

[11] Long Tail Analysis: https://threatpost.com/long-tail-analysis-hope-cybercrime-battle/155992/

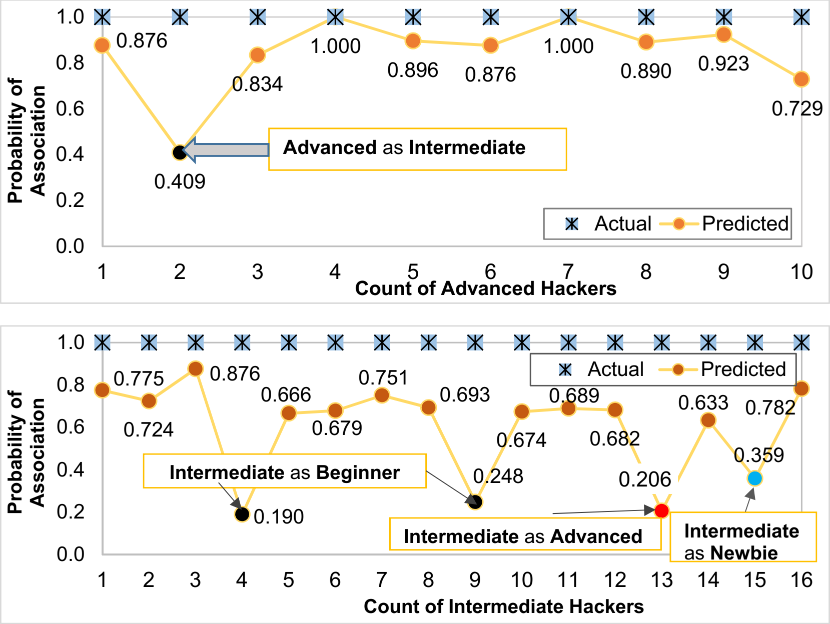


Fig. 3. (a) “Actual” versus “Predicted” probability for *Advanced* hackers (with test data).
(b) “Actual” versus “Predicted” probability for *Intermediate* hackers (with test data).

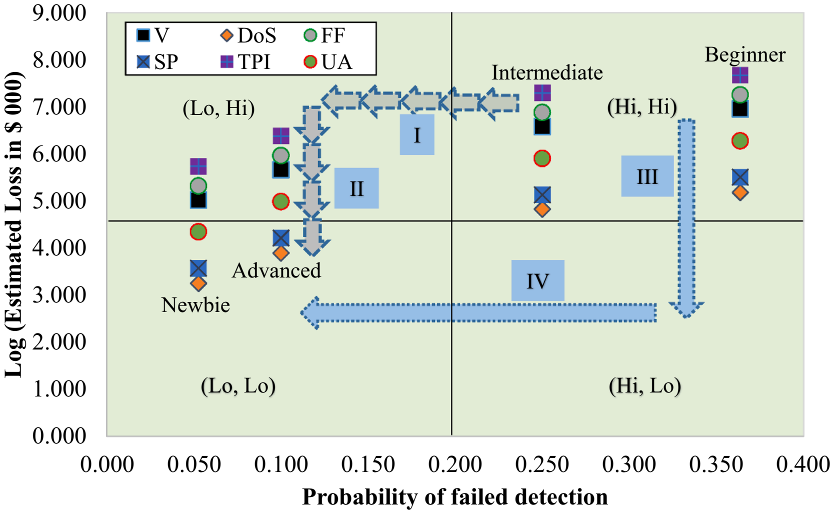


Fig. 4. Strategy map for recovery of a firm after suffering cyber-attacks.

attacks.[12] Similarly, hacker Anna Senpai launched the Mirai botnet attacks, who later distributed the source codes in *hackforums.net*,[13] facilitating malware replication by other novice hackers. However, sharing attachments and exploits in a forum by hackers has likely changed over time and commensurate with the continued advances in IT security technology.

Next, we examined the role of *sentiments* (both positive and negative) in determining the expertise level of hackers in dark forums. Additionally, we noted that both positive ($\beta = 0.132^{**}$; $\beta = 0.610^{**}$; $\beta = 0.524^{**}$) and negative ($\beta = 0.056^{**}$; $\beta = 0.341^{**}$; $\beta = 0.284^{**}$) sentiments emerged as significant and highly effective predictors of expertise, especially with the *intermediate* and *advanced* hackers. We build our corpus of cyber keywords during the analysis and assigned relative weights before calculating the sentiment scores [21]. *Advanced* hackers show the highest opinion values in their messages and discussions, followed by *beginners* and *newbies*. Our findings are novel among the recent literature on expert identification in hacker forums, especially where Benjamin et al. [6] have encouraged future scholars to conduct a sentiment-based content analysis of conversations for gaining insights into dark forums. In addition, Li et al. [31] have identified key sellers using sentiment analysis of customer reviews across carding communities in the dark web. Findings from our study are also congruent with healthcare forums [13,14], electronic marketplaces [9,52], stakeholder analysis [28], and online product reviews on Amazon [30,43].

Then, we find that *the presence of cybersecurity keywords* in the messages strongly links to all levels of expertise in hacker forums ($\beta = 0.734^{***}$; $\beta = 1.650^{***}$; $\beta = 1.512^{***}$). Additionally, we note that the usage of “cybersecurity” keywords is much higher for *advanced* and

[12] Home Depot Hit By Malware: https://krebsonsecurity.com/2014/09/home-depot-hit-by-same-malware-as-target/

[13] Who is Anna-Senpai? https://krebsonsecurity.com/2017/01/who-is-anna-senpai-the-mirai-worm-author/

intermediate hackers than *beginners* and *novices*. Our study also reinforces that it is the most relevant predictor for recognizing expert hackers in dark forums (see Table 6 and Table 10). Besides, this finding is supported by the presence of the Pareto effect among key hackers while replying to technical queries in dark forums [8] and underground carding shops [31]. Our findings partly relate to extant studies examining the presence of keywords such as (i) *attack vectors* (e.g. XSS, DDoS) and *hacking concepts* (e.g. shellcode) [6]; (ii) *knowledge-providing* and *knowledge-acquisition* keywords [27,36,55]. Recent studies have reported similar behaviour among key actors, such as professionals in online healthcare communities [14,38] and knowledge forums [29].

Finally, we examine the influence of special characters on the expertise level of a hacker. While a few current studies highlight the use of inimitable symbols within the messages such as excess punctuations, URLs, italics and font colours to create diversity [6], or when the hacker uses emoticons, smileys, and non-alphanumeric patterns in their message [8], our study did not report any significant effect on the expertise in hacker forums ($\beta = -0.325$; $\beta = -0.227$; $\beta = -0.333$). Possible explanations could be as follows. First, hackers do not follow the standard semantics of the English language while communicating on dark forums. Because of the same reason, we did not adopt *text readability* as a predictor of expert hackers in this study. Second, hackers might be more interested in exchanging malicious knowledge, assets such as exploits and tutorials, through URLs and secure file transfer mechanisms [8].

6.2. Core incremental contributions to IS research

Our study has several theoretical contributions, mainly towards *Community of Practice Theory* [31] and *Social Exchange Theory* [10] in the context of the "dark side of IT" and cybersecurity analytics. First, this study extends the following aspects of *Community of Practice Theory:* (i) Identifies the presence of cognitive learning behaviour within the dark web communities, especially hacker forums. In particular, this study identifies the acquisition mechanism of malicious knowledge for a hacker that is moderately or sometimes entirely dependent on expert peers. For instance, beginners can learn from advanced hackers how to create attacks using exploits through the discussion boards in these dark forums. (ii) Reinforces the attributes of this learning mechanism and highlights its development through interactions, experiences, and external media influences such as sharing of attachments where top hackers often distribute video tutorials on cyber-attacks and source codes. While there are academic studies that reveal top hackers using the embedded social networks within these dark forums, they are yet to acknowledge hacker communities as hotspots for cognitive enhancement and social learning. In this manner, this study re-establishes the following attributes of a community-of-practice in the context of dark forums: *problem-solving*, *presence of a set of focal problems* or *passion about a topic*, *request for information*, *reuse of assets*, *improvement of social capital*, and finally *generation of artefacts.*[14] This is the primary incremental contribution to the existing IS research on the "dark side of IT" and cybersecurity analytics.

Second, this study draws on Social Exchange Theory and contributes in the following ways: (i) integrates a set of robust and unique features into the subsequent cost-benefit analysis through a knowledge exchange after the "social learning." This study employed *forum-usage* features such as *count of thread participation*, *message replies per thread*, *initiation of discussion boards*, and *message-content* features such as *cybersecurity keywords*, *positive* and *negative sentiment content* of a message, *sharing of attachments*. (2) Senior hackers establish their ranks such as "advanced hackers" or "intermediate hackers" by disseminating their accumulated "knowledge" capital by responding to queries and doubts posed by junior hackers (e.g. Pareto effect among key hackers while replying to technical queries [8]). In the "social exchange", elite hackers enjoy intangible rewards (i.e. sense of achievement and pride) and trust among peers. This is the second contribution to the existing literature on cybersecurity analytics and dark forums.

6.3. Lessons learnt for IT and business executives

The managerial contributions of this study include the following. First, this study identifies that the expertise level of hackers in a dark forum can be revealed by analyzing their *forum usage* and *message-content* behaviour. Among *forum usage* predictors: *replies per thread* and *discussion threads initiated* by hackers are significant, while *usage of cybersecurity keywords*, *attachments*, and *sentiment content* are the top *message-content* predictors. Such findings have interesting ramifications for organizations seeking to study what these malicious actors are discussing and, in turn, help identify who they are. Cybersecurity analysts can design proactive tools (e.g. analytics dashboard) where these predictors can be employed to generate visual insights based on the discussion boards in these dark forums.[15] For instance, after analysts detect a hacker "HttP-NuKe" to be significant across many discussions in the *Hackhound* forum, the next steps will be to (i) monitor all subsequent discussion activities of "HttP-NuKe" within this forum, as well as other dark forums and social media groups that he/she is part of; (ii) Next, they can extract the top-ranked keywords[16] to track the attack-vector and the possible victim organization(s). For instance, if the keywords "SQL injection",[17] ".gov", ".edu" appear significant, analysts will realize that web applications hosted within the U.S. government agencies and universities are the potential targets. Recently, in December 2020 and January 2021, hackers exploited Accellion's secure file transfer appliance (FTA) using SQL injection attacks,[18] whose victims were the University of Colorado and the State of Washington.[19]

Second, this study proposes a novel risk-evaluation metric {expected loss, probability, attack-type, hacker expertise} that organizations can apply to gauge different expertise levels of hackers who can launch various types of cyber-attacks (e.g. virus attacks, DoS, and financial fraud attacks). Besides, they can classify attackers across two dimensions on a 2-by-2 strategy map: *expected loss* that they can inflict and the *likelihood of failed detection* even after executing cyber-threat intelligence and analytics (see Fig. 4).

Third, firms can now use analytics for better decision-making, risk mitigation and allocation of IT security budgets, depending on the type of attack and the attackers involved. Our framework performs well for *advanced* and *newbies*, who are easily detected than midway groups, i.e. *beginners* and *intermediates.* Therefore, according to our proposed mitigation strategy-map (see Fig. 4), these hackers can run undetected and launch DoS, DDoS and system penetration attacks, thereby placing them in the (Hi, Hi) zone on our strategy map. The mitigation steps for such attacks are given as *Step I*: firms can decrease the rate of misdetection through technological improvement in the classifier or by refining the rule-set if necessary. This step will bring down the probability to the (Lo, Hi) zone. *Step II:* firms can reduce financial impact by internal mechanisms such as revision of existing cybersecurity policy, employee training [18,19], or resort to external risk-transfer mechanisms such as third-party cyber-insurance [39,40]. This step will bring down the expected loss to the (Lo, Lo) zone.

[14] Introduction to communities of practice: https://wenger-trayner.com/introduction-to-communities-of-practice/

[15] Hackhound Dashboard: https://public.tableau.com/app/profile/sagarsamtani07/viz/HackerForums/Hackhound

[16] We could calculate top keywords more accurately with the proposed overlap score instead of Samtani et al. [45]

[17] SQL Injection: https://us-cert.cisa.gov/security-publications/sql-injection

[18] FireEye: https://www.accellion.com/sites/default/files/trust-center/accellion-fta-attack-mandiant-report-full.pdf

[19] The Accellion Breach Keeps Getting Worse: https://www.wired.com/story/accellion-breach-victims-extortion/

Next, if these hackers execute phishing attacks that steal personal information from individuals [15], this study places them in the (Hi, Hi) zone of the strategy map. Typically, these attacks have the highest financial impact among all types of cyber-attacks.[20] To address these attacks, our framework recommends the mitigation steps as *Step III*: firms need to train employees to understand phishing emails and URLs based on their susceptibility and familiarity of source [15] because the variants of phishing attacks change fast [3].[21] These drives will endorse cyber awareness [7,18] and improve cyber-hygiene [49] in an organization. This step will bring the expected loss to the (Hi, Lo) zone. (ii) *Step IV*: firms can decrease the rate of misdetection through additional investments in IT security and anti-phishing filters. This step will bring down the probability to the (Lo, Lo) zone.

6.4. Concluding remarks

This study presented a two-stage framework to identify the multiple levels of hacker expertise within a dark forum using a multi-class identification technique. The inaccurate detection of hacker expertise led to a firm's expected loss. This study provides an important contrast to the current literature on the “dark side of IT” and cybersecurity analytics, where standard classification algorithms are employed to find significant discussion themes using topic-modelling, SNA, and sentiment-analysis only. To address these gaps, we presented a new direction of modelling cyber risk assessment and mitigation through analysis of hacker messages in dark forums so that organizations can evaluate potential hackers without actual interaction. This study could be extended further as follows. First, scholars can perform text mining and classification analysis with multiple forums in languages other than English and subsequently perform risk-mitigation exercises. Second, future scholars can perform an in-depth longitudinal examination of hacker messages for each hacker. In this way, any significant change or evolution in the levels of hacker expertise can be traced across the period of analysis and then derive novel insights.

Acknowledgements

The authors express their sincerest thanks to the anonymous referees for their constructive suggestions. The authors also thank the Editor-in-Chief of this journal, Professor James R. Marsden, for his overall inputs and support throughout the revision process.

References

[1] H. Akman, C. Plewa, J. Conduit, Co-creating value in online innovation communities, Eur. J. Mark. 53 (6) (2019) 1205–1233.

[2] A. Armstrong, J. Hagel, Net Gain: Expanding Markets through Virtual Communities, Harvard Business School, 1997.

[3] N. Azeez, S. Misra, I.A. Margaret, L. Fernandez-Sanz, Adopting automated whitelist approach for detecting phishing attacks, Comp. Security (2021) 102328.

[4] V. Benjamin, H. Chen, Securing cyberspace: identifying key actors in hacker communities, in: IEEE International Conference on Intelligence and Security Informatics, 2012, June, pp. 24–29.

[5] V. Benjamin, B. Zhang, J.F. Nunamaker, H. Chen, Examining Hacker participation length in cybercriminal internet-relay-chat communities, J. Manag. Inf. Syst. 33 (2) (2016) 482–510.

[6] V. Benjamin, J.S. Valacich, H. Chen, DICE-E: a framework for conducting Darknet identification, collection, evaluation with ethics, MIS Q. 43 (1) (2019).

[7] B. Biswas, A. Mukhopadhyay, G-RAM framework for software risk assessment and mitigation strategies in organizations, J. Enterp. Inf. Manag. 31 (2) (2018) 276–299.

[8] B. Biswas, A. Mukhopadhyay, G. Gupta, “Leadership in action: how top hackers behave” a big-data approach with text-mining and sentiment analysis, in: Proceedings of the 51st Hawaii International Conference on System Sciences, 2018.

[9] B. Biswas, P. Sengupta, D. Chatterjee, Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques, Decis. Support. Syst. 135 (2020) 113324.

[10] P. Blau, Exchange and Power in Social Life, John Wiley & Sons, New York, 1964.

[11] M. Bouguessa, L.B. Romdhane, Identifying authorities in online communities, ACM Trans. Intell. Syst. Technol. (TIST) 6 (3) (2015) 30–52.

[12] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[13] L. Chen, A. Baird, D. Straub, Why do participants continue to contribute? Evaluation of usefulness voting and commenting motivational affordances within an online knowledge community, Decis. Support. Syst. 118 (2019) 21–32.

[14] L. Chen, A. Baird, D. Straub, A linguistic signaling model of social support exchange in online health communities, Decis. Support. Syst. 130 (2020) 113233.

[15] R. Chen, J. Gaia, H.R. Rao, An examination of the effect of recent phishing encounters on phishing susceptibility, Decis. Support. Syst. 133 (2020) 113287.

[16] X. Cheng, L. Su, X. Luo, J. Benitez, S. Cai, The good, the bad, and the ugly: impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing, Eur. J. Inf. Syst. (2021) 1–25.

[18] E. Dincelli, I. Chengalur-Smith, Choose your own training adventure: designing a gamified SETA artefact for improving information security and privacy through interactive storytelling, Eur. J. Inf. Syst. 29 (6) (2020) 669–687.

[19] Y.H. Fang, C.Y. Li, Leveraging sociability for trust building on social commerce sites, Electron. Commer. Res. Appl. 40 (2020) 100907.

[20] S. Faraj, S. Kudaravalli, M. Wasko, Leading collaboration in online communities, MIS Q. 39 (2) (2015) 393–412.

[21] E. Fersini, E. Messina, F.A. Pozzi, Expressive signals in social media languages to improve polarity detection, Inf. Process. Manag. 52 (1) (2016) 20–35.

[22] J. Grisham, S. Samtani, M. Patton, H. Chen, Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence, in: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), 2017, July, pp. 13–18.

[23] J. Hacker, R. Bernsmann, K. Riemer, Dimensions of user behavior in Enterprise social networks, in: Social Knowledge Management in Action, Springer, 2017, pp. 125–146.

[24] J. Han, J. Pei, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, 2011.

[25] G.C. Homans, Social behavior as exchange, Am. J. Sociol. 63 (6) (1958) 597–606.

[26] S.M. Horng, C.L. Wu, How behaviors on social network sites and online social capital influence social commerce intentions, Inf. Manag. 57 (2) (2020) 103176.

[27] C. Huang, Y. Guo, W. Guo, Y. Li, HackerRank: identifying key hackers in underground forums, Int. J. Distrib. Sens. Networks 17 (5) (2021), 15501477211015145.

[28] S. Jiang, H. Chen, J.F. Nunamaker, D. Zimbra, Analyzing firm-specific social media and market: a stakeholder-based event analysis framework, Decis. Support. Syst. 67 (2014) 30–39.

[29] J. Jin, Y. Li, X. Zhong, L. Zhai, Why users contribute knowledge to online communities: an empirical study of an online social Q&a community, Inf. Manag. 52 (7) (2015) 840–849.

[30] S.J. Kim, E. Maslowska, A. Tamaddoni, The paradox of (dis) trust in sponsorship disclosure: the characteristics and effects of sponsored online consumer reviews, Decis. Support. Syst. 116 (2019) 114–124.

[31] J. Lave, E. Wenger, Learning in doing: Social, cognitive, and computational perspectives, in: Situated Learning: Legitimate Peripheral Participation, Press, Cambridge University, 1991.

[33] H.C. Lin, C.M. Chang, What motivates health information exchange in social media? The roles of the social cognitive theory and perceived interactivity, Inf. Manag. 55 (6) (2018) 771–780.

[34] R. McMillan, Threat Intelligence, Available at: https://www.gartner.com/doc/2487216/definition-threat-intelligence, 2013.

[35] C.D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, 1 (1), 496, Cambridge University Press, Cambridge, 2008.

[36] E. Marin, J. Shakarian, P. Shakarian, Mining key-hackers on darkweb forums, in: 2018 1st International Conference on Data Intelligence and Security (ICDIS), IEEE, 2018, April, pp. 73–80.

[37] S.A. Matei, A.A. Jabal, E. Bertino, Do sticky elites produce online knowledge of higher quality?, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2017, pp. 72–79.

[38] R. Mousavi, T.S. Raghu, K. Frey, Harnessing artificial intelligence to improve the quality of answers in online question-answering health forums, J. Manag. Inf. Syst. 37 (4) (2020) 1073–1098.

[39] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, S.K. Sadhukhan, Cyber-risk decision models: to insure IT or not? Decis. Support. Syst. 56 (2013) 11–26.

[40] A. Mukhopadhyay, S. Chatterjee, K.K. Bagchi, P.J. Kirs, G.K. Shukla, Cyber risk assessment and mitigation (CRAM) framework using logit and Probit models for cyber insurance, Inf. Syst. Front. 21 (5) (2019) 997–1018.

[41] I. Park, S. Sarnikar, J. Cho, Disentangling the effects of efficacy-facilitating informational support on health resilience in online health communities based on phrase-level text analysis, Inf. Manag. 57 (8) (2020) 103372.

[42] J.W. Pennebaker, R.L. Boyd, K. Jordan, K. Blackburn, The Development and Psychometric Properties of LIWC2015, 2015.

[43] M. Salehan, D.J. Kim, Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics, Decis. Support. Syst. 81 (2016) 30–40.

[44] S. Samtani, H. Chen, Using social network analysis to identify key hackers for keylogging tools in hacker forums, in: 2016 IEEE Conference on Intelligence and Security Informatics (ISI), IEEE, September, 2016, pp. 319–321.

[20] FBI Internet Crime Complaint Center Report (2020): https://bit.ly/34LCzlL

[21] Phishing and social engineered attacks remain the most popular category of attack in recent years. These attack techniques change very fast e.g. COVID phishing frauds and anti-phishing filters have trouble to detect new variants.

[45] S. Samtani, R. Chinn, H. Chen, J.F. Nunamaker Jr., Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence, J. Manag. Inf. Syst. 34 (4) (2017) 1023–1053.
[46] S. Samtani, Y. Chai, H. Chen, Linking exploits from the dark web to known vulnerabilities for proactive cyber threat intelligence: an attention-based deep structured semantic model, MIS Q. (2021), https://doi.org/10.25300/MISQ/2022/15392 (forthcoming).
[47] M. Siering, J.A. Koch, A.V. Deokar, Detecting fraudulent behavior on crowdfunding platforms: the role of linguistic and content-based cues in static and dynamic contexts, J. Manag. Inf. Syst. 33 (2) (2016) 421–455.
[48] S.L. Vargo, R.F. Lusch, Service-dominant logic: continuing the evolution, J. Acad. Mark. Sci. 36 (1) (2008) 1–10.
[49] A. Vishwanath, L.S. Neo, P. Goh, S. Lee, M. Khader, G. Ong, J. Chin, Cyber hygiene: the concept, its measure, and its initial tests, Decis. Support. Syst. 128 (2020) 113160.
[50] M.M. Wasko, S. Faraj, Why should I share? Examining social capital and knowledge contribution in electronic networks of practice, MIS Q. (2005) 35–57.
[51] E.C. Wenger, W.M. Snyder, Communities of practice: the organizational frontier, Harv. Bus. Rev. 78 (1) (2000) 139–146.
[52] J. Wu, J. Cai, X.R. Luo, J. Benitez, How to increase customer repeated bookings in the short-term room rental market? A large-scale granular data investigation, Decis. Support. Syst. 143 (2021) 113495.
[53] K. Xie, Y. Wu, J. Xiao, Q. Hu, Value co-creation between firms and customers: the role of big data-based cooperative assets, Inf. Manag. 53 (8) (2016) 1034–1048.
[54] X. Yang, G. Yang, J. Wu, Y. Dang, W. Fan, Modeling relationships between retail prices and consumer reviews: a machine discovery approach and comprehensive evaluations, Decis. Support. Syst. 145 (2021) 113536.
[55] X. Zhang, A. Tsang, W.T. Yue, M. Chau, The classification of hackers by knowledge exchange behaviors, Inf. Syst. Front. 17 (6) (2015) 1239–1251.

Baidyanath Biswas is an Assistant Professor of MIS and Analytics Group at the International Management Institute Kolkata, India. His research has appeared in *Decision Support Systems, Electronic Markets, Computers in Industrial Engineering,* and the Journal of Enterprise Information Management. Baidyanath is also associated with top peer-reviewed international conferences, namely, HICSS and ICIS. He has a rich industry-experience of nine years working as a mainframe and DB2 database analyst at Infosys and IBM. Currently, Baidyanath serves as the Associate Editor of the Global Business Review journal.

Arunabha Mukhopadhyay is a Professor of Information Technology & Systems Area at Indian Institute of Management Lucknow (IIM Lucknow). He has obtained his Ph.D. and Post Graduate Diploma in Business Management (PGDBM) from the Indian Institute of Management Calcutta (IIM Calcutta), in the area of Management Information Systems. He was awarded the Infosys scholarship during his Ph.D. He has published in various referred journals and conferences including *Decision Support Systems (DSS) Information Systems Frontier (ISF), Journal of Global Information Technology Management (JGITM), JIPS, International Journal of Information Systems and Change Management (IJISCM), Decision, IIMB Review, Hawaii International Conference on System Sciences (HICSS), Americas Conference on Information Systems (AMCIS), Pre-International Conference On Information Systems (ICIS) workshops,* etc. He is the recipient of the Best Teacher in Information Technology Management in 2013 and 2011, by Star-DNA group B-School Award and 19th Dewang Mehta Business School Award, in India, respectively.

Sudip Bhattacharjee is a Professor in the School of Business, University of Connecticut. He currently serves as Senior Research Fellow, US Census Bureau. Previously, he served as Chief, Center for Big Data Research and Applications, US Census Bureau. He is a Visiting Faculty at EMLYON Business School, France, and Indian School of Business. He was a Visiting Professor at GE Global Research Center, USA. He has previously served as the Assistant Dept. Head of Operations and Information Management, and as the Executive Director of MBA Programs, both in the School of Business, University of Connecticut. His research interests include data driven IT and operations management and policy, information systems economics, energy informatics, digital goods and markets, and closed loop supply chains. His research has appeared in premier journals such as *Management Science, INFORMS Journal on Computing, Journal of Business, Journal of Law and Economics, ACM Transactions, Journal of Management Information Systems, IEEE Transactions,* and other leading peer-reviewed publications. He serves or has served as Associate Editor for *Information Systems Research* (for 5 years), Special Issue Editor for *ACM Transactions on Management Information Systems*, guest AE for *MIS Quarterly* and *Decision Sciences Journal.*

Ajay Kumar is an Assistant Professor at the AIM Research Center on Artificial Intellegence in Value Creation, EMLYON Business School in France. His research and teaching interests are in data and text mining, decision support systems, business intelligence and enterprise modelling. He has been Postdoctoral Fellow at Massachusetts Institute of Technology and Harvard Business School. He has published several research papers in reputed journals, including *International Journal of Production Economics*, *Industrial Marketing Management*, *Telematics & Informatics*, *Technological Forecasting & Social Change*, *Annals of Operation Research*, *International Journal of Production Research*, etc.

Dursun Delen is the holder of Spears and Patterson Endowed Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. He authored/co-authored 100+ journal and 40+ peer-reviewed conference proceeding articles. His research has appeared in major journals including *Decision Sciences, Journal of Production Operations Management, Decision Support Systems, Communications of the ACM, Computers and Operations Research, Computers in Industry, Artificial Intelligence in Medicine, International Journal of Medical Informatics, Expert Systems*, among others. He has recently published ten books/textbooks in the broad area of Business Intelligence and Business Analytics. He is often invited to national and international conferences and symposiums for keynote addresses, and companies and government agencies for consultancy/education projects on Analytics related topics. He is currently serving as the editor-in-chief, senior editor, associate editor, and editorial board member of more than a dozen academic journals.