Computer Communications 179 (2021) 285–295

Contents lists available at ScienceDirect

Computer Communications

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

Using data mining techniques to explore security issues in smart living environments in Twitter

Jose Ramon Saura [a,*], Daniel Palacios-Marqués [b], Domingo Ribeiro-Soriano [c]

[a] Rey Juan Carlos University, Madrid, Spain
[b] Universitat Politècnica de València, Valencia, Spain
[c] University of Alcalá, Madrid, Spain

ARTICLE INFO

Keywords:
Home assistant
IoT
Sentiment analysis
Data mining
Twitter
UGC

ABSTRACT

In present-day in consumers' homes, there are millions of Internet-connected devices that are known to jointly represent the Internet of Things (IoT). The development of the IoT industry has led to the emergence of connected devices and home assistants that create smart living environments. However, the continuously generated data accumulated by these connected devices create security issues and raise user's privacy concerns. The present study aims to explore the main security issues in smart living environments using data mining techniques. To this end, we applied a three-sentence data mining analysis of 938,258 tweets collected from Twitter under the user-generated data (UGD) framework. First, sentiment analysis was applied using Textblob which was tested with support vector classifier, multinomial naïve bayes, logistic regression, and random forest classifier; as a result, the analyzed tweets were divided into those expressing positive, negative, and neutral sentiment. Next, a Latent Dirichlet Allocation (LDA) algorithm was applied to divide the sample into topics related to security issues in smart living environments. Finally, the insights were extracted by applying a textual analysis process in Python validated with the analysis of frequency and weighted percentage variables and calculating the statistical measure known as mutual information (MI) to analyze the identified n-grams (unigrams and bigrams). As a result of the research 10 topics were identified in which we found that the main security issues are malware, cybersecurity attacks, data storing vulnerabilities, the use of testing software in IoT, and possible leaks due to the lack of user experience. We discussed different circumstances and causes that may affect user security and privacy when using IoT devices and emphasized the importance of UGC in the processing of personal data of IoT device users.

1. Introduction

The rapid development of information technologies, the Internet, and artificial intelligence, in all their versatile forms, has allowed the development of connected ecosystems known as smart living environments [1]. In today's connected era, there are many sources of information that continuously generate data on user habits, and these data generate concerns about data privacy risks [2]. This innovative paradigm [3], linked to the IoT sector, has given rise to the concept of IoE (Internet of Everything), which links new theoretical frameworks for the study of emerging research areas related to privacy and data security [4], as well as the usability of these new technologies [5] for various applications on the industrial, professional, or personal level [6].

In the last decade, connected devices and IoT have been among the most rapidly growing industries [7]. Therefore, millions of devices are connected to the Internet in consumers' homes [8] to facilitate users' performance of daily tasks. Today, many consumers actively use technology in their lives and develop their business in connected ecosystems [9]. Accordingly, Bouncken and Barwinski [10] highlighted the importance of the shared digital identity concept as a collective self-concept of an in-group towards the creation, application, development, and emergence of digital technology that creates a sense of community. The use of this technology can encourage the feeling of both security and ease of use of these devices [11]. However, users may not be aware of the privacy and personal information breaches that can occur if these data are managed illegally or unfairly by technological companies or by third parties [5,12].

Accordingly, several previous studies warned about the potential risks of the connected devices' use when data privacy settings are not configured or are configured in a wrong way [13,14]. In general, data resulting from user actions requested by the devices can be used to identify patterns, trends, or predict user behaviors [15].

Smart living environments are equipped with numerous devices that collect information in the form of commands and requests issued by

* Corresponding author.
E-mail addresses: joseramon.saura@urjc.es (J.R. Saura), dapamar@doe.upv.es (D. Palacios-Marqués), ribeders@uv.es (D. Ribeiro-Soriano).

https://doi.org/10.1016/j.comcom.2021.08.021
Received 16 January 2021; Received in revised form 31 July 2021; Accepted 24 August 2021
Available online 2 September 2021

users [7,16]. Based on these commands, these devices can perform specific functions, such as turning the light on or off, playing music, answering questions, turning the TV or other connected machines and devices on or off, organizing the calendar, the agenda, or sending e-mails [17]. In the literature, these connected homes are collectively called smart living environments [18]. These environments have been studied from the theoretical [19], security [20], innovation [21], and future [22] perspectives.

Importantly, even if users make only limited use of these technologies to digitalize their homes or make part of a connected community [23], they may be unaware that the Internet-connected devices in their smart living environments can generate data about their personal habits and interests, which could violate their privacy [24]. In this respect, Lin and Bergmann [25] and Chow [26] highlighted the concerns about user privacy and security of users, which is an important topic to thoroughly explore in the future [27].

It is important to note that, to date, several contributions analyzed how IoT sensors are added to the home networks when connecting new devices to the Internet [28]; similarly, other studies investigated the ways to make these devices safer from the technical point of view [29]. Overall, connected devices are linked to platforms that manage the data [30], which implies that security of these cloud-based devices should be a priority [31].

However, at present, most Internet-based connections in homes go through the same channel—namely, the router that provides Internet connection [32]. Furthermore, when purchasing connected devices for their homes, consumers focus on connecting them to other devices, such as TVs, smart lightbulbs, smart toothbrushes, smart coffee mugs, smart ovens, smart juicers, smart thermostats, and so forth, overlooking thus the importance of improving device privacy [33]. Accordingly, several authors argued that consumers are the party responsible for safeguarding their own assistants that go through the router, and the lack of knowledge about the importance of performing appropriate IoT devices installations in smart home environments can make user information vulnerable [34,35].

In principle, users should be aware what kind of information about them can be used when they connect their devices [36], as well as understand that data breaches, which can put their privacy at risk, may occur [37]. However, although the main wizards and IoT devices are developed by large technology companies (e.g., Google, Amazon, or Apple) that, in principle, invest much effort into making their devices secure, in reality, the multitude of accesses make these devices unsafe [38].

Furthermore, while some authors talk about achieving standardization to protect IoT devices in smart living environments [39], in reality, this technology is only developing and must be further improved as its efficiency and effectiveness progress [40]. An important issue in this respect is educating users about potential exposure and/or violation of their personal data in the event of computer attacks [41]. In addition, the speed at which new IoT device systems develop and grow on the market suggests that 'security first' is a concept that has not yet been effectively applied [13].

Overall, new connected devices are a result of constant innovation [21]. Innovation drives the fastest possible launch of new devices in the market, and such launches frequently occur without rigorous testing against possible attacks and security beaches [15]. Compromising on security of new devices leads to the quick detection of security gaps and to the creation of patches [42]. Over time, as more gaps get detected and fixed, the devices will start to become more secure [43]. However, this improvement can be counted on only when the devices are already being used by consumers in the market [44]. Therefore, the application of 'security first' during the development lifecycle of IoT products should be an essential point of companies' agenda [44].

While privacy is a complex issue, users are growing increasingly aware that connected devices must follow certain security protocols [42]. However, privacy and security of these devices remain a challenge for users who make standard use of these products in their smart living environments [45].

Therefore, in order to overcome these challenges of security issues concerns in smart living environments, we aim to explore the security issues of IoT devices and home assistant in smart living environments [25] using user generated-content (UGC) and user generated-data (UGD) published in Twitter as a main source of data. Our second goal is to establish how these security issues can affect user privacy [46]. Specifically, the present study addresses the following two research questions:

- What are the main security issues of IoT devices and home assistants in smart living environments?
- How can security issues of IoT devices and home assistants in smart living environments affect user privacy?

The novelty and originality of the present study lies in that, in order to answer the aforementioned research questions, we used a novel exploratory methodology based on extracting insights related to security and privacy of smart living environments, as this approach has not been used in the field before using Twitter as a source of UGC and UGD. Furthermore, to reinforce the results and to allow for the use of our findings in future research on smart living environments using the frameworks of UGC and UGD in Twitter, we have constructed validation of our results using Computer-Aided Text Analysis (CATA) theoretical framework.

In terms of the methodology, we first applied sentiment analysis using Textblob which was tested with support vector classifier, multinomial naïve Bayes, logistic regression, and random forest classifier to divide the sample into tweets expressing positive, negative, and neutral sentiment. Next, a Latent Dirichlet allocation (LDA) algorithm was applied to divide the sample into themes related to security issues. The topics were then tested with the keyness and *p*-value metrics to measure their relevance in the database. Finally, insights were extracted by applying a textual analysis process in Python validated with the analysis of count and weighted percentage variables following the CATA validation rules proposed by Short et al. [47] and Täuscher et al. [48]. Finally, we analyzed the n-grams (unigrams and bigrams) corresponding to the content of the database extracted from Twitter.

The remainder of this paper is structured as follows. Section 2 presents the literature review Section 3 provides further detail on the methodology used in the present study. The results are reported in Section 4. The paper concludes with a discussion (Section 5) and conclusions (Section 6), including an outline of theoretical and practical contributions, as well as limitations and further research directions.

2. Literature review

Nowadays, people share massive amounts of personal information, as well as their thoughts or opinions, through social networks [49]. Social network platforms such as Twitter enable users to interact with each other through the generation and sharing of content [50], which allows one to obtain significant insights into user opinions. Moreover, this content is used by enterprises to improve their managerial decisions, as well as online marketing and communication strategies. In the form of comments, opinions, and so forth, the content published on social networks such as Twitter is analyzed under the framework known as UGC or UGD. The UGC has been consolidated as a relevant approach for the identification of insights into emerging events in social networks in the last decade [51].

In this respect, as argued by Sarlan et al. [52], Twitter is one of the social media that is gaining popularity in terms of spreading public and private opinions. Therefore, sentiment analysis can be applied to computationally measure customers' perceptions. The study of these perceptions and opinions can help companies to improve their business strategies or identify pertinent issues in different areas [53].

With regard to smart environment applications (SEA) that are directly linked to smart home environments, Alam et al. [54] analyzed how social media mining algorithm can help mine user opinions by clustering textual data into different behaviors. This method combined with sentiment analysis and a neural network-driven approach was demonstrated to improve SCAs' performance. Furthermore, Webberley et al. [55] argued that the level of retweeting of a tweet determines its interestingness, as it could be seen the signal used for affective simulation. Therefore, the aim of this study was to help address the "filter bubble" problem by revealing interesting content.

Today, users interact with a wide range of IoT devices, ranging from sensors people carry on their wrists to network-connected thermostats [56]. Likewise, users also extensively use home assistant devices to tidy up rooms, sort laundry, and manage storage when owners are away [57]. In this context, it becomes increasingly difficult to preserve user data privacy and control how data are treated [58].

Specifically, while smartphone users manage the permissions granted to the apps they deploy on their devices, IoT users interact with technologies they did not install and are hardly ever aware of possible privacy threats [59]. Unaware of the risks of IoT systems, people give private information about their meaningful choices [58], which makes them vulnerable through hardware, software, and side-channels, and the risks are exacerbated when combined.

In this connection, Alrawi et al. [60] has recently argued that the IoT devices rely on insecure protocols that do not offer confidentiality or integrity and, therefore, may lack endpoint verification. In another study, home assistants were found to enforce an endpoint to verify the cloud identity [10], but not to enforce the verification of the application identity. This leads to vulnerable endpoints when replayed requests are intended for other applications [61]. Therefore, in the event of an attack, the information provided by home assistants may reveal a highly sensitive private information about user lifestyles [62].

Another group of concerns is related to the cloud context where sensitive data are stored [63]. These concerns and opinions could be seen on Twitter UGC where the interaction over the past few years has increased.

Indeed, social media platforms have changed how information is produced and consumed [64]. In a study on that applied sentiment analysis to analyze conferences, Parra et al. [65] found that well-established communities in Twitter present stable patterns, and human–computer interaction conferences show consistently more emotion and a higher number of positive tweets than conferences in analytical areas [65]. This evidence validates the use of this social network in research on emerging causes [62].

A summary of relevant studies that used social networks to extract IoT-related insights using UGC or UGD or highlighted the relevance of security issues in IoT devices and home assistants is provided in Table 1.

3. Methodology

As mentioned above, in the present study, we used a three-step methodology to extract UGC insights from Twitter. Data-mining processes are generally used to create knowledge and identify indicators to explore and enrich the research [64]. In the present study, we used the CATA theoretical framework [47]. CATA defines potential approaches to construct validation using computer-aided text analysis tools and to capture theoretically-based constructs of interest. As indicated by Short et al. [47], Pollach [71], and McKenny et al. [72], software programs appropriate for studies based on CATA include VBPro, CATPAC, Concordance, DICTION, General Inquirer, LIWC, NVivo, and MECA.

Furthermore, as argued by Xuanyang et al. [73] and Leon [74], in order to improve both the results and the validity of both knowledge and theory of the studies based on CATA, the computer text analysis approach, as well as technologies such as data mining, machine learning, or prediction algorithms, should be used. Krippendorff [75] also considered the use of data mining, machine learning algorithms and other artificial intelligence-based approaches to improve the identification of insights using computer text analysis methods, highlighting that these approaches are more efficient than human coding in terms of cost and speed.

Accordingly, in the present study, we performed sentiment analysis using Textblob. This approach can be used with technological tools and approaches such as Support Vector Machines (SVM), Support Vector Classifiers (SVC), Naïve Bayes or Logistic Regression, among others [76]. In the present study, we used Textblob, a classification tool developed in Python, which was tested with a support vector classifier, multinomial naïve Bayes, logistic regression, and random forest classifier [77].

Secondly, an LDA algorithm – a topic-modeling tool – was applied to divide the sample into topics related to security issues in smart living environments and tested with keyness and *p*-value metrics. Both topic modeling and the use of LDA algorithms include content analysis and development of knowledge extraction models [78]. Similarly, the combination of the sentiment analysis results with topic modeling increases the relevance of the procedure validation with CATA, as the results are filtered by different combinations of text analysis and debugging methods [79]. LDA is widely used in scientific research and content analysis approaches [75,78].

Thirdly, as indicated by Short et al. [47], insights can be extracted by applying the textual analysis process. This process can be developed using tools such as NVivo or programming languages such as Python or R [73,74]. These approaches are used for the analysis of language dictionaries to calculate the numbers of specific keywords, the weighted percentage, or n-grams indicators. These procedures were also used in the present study, because the textual analysis developed in Python is structured in theoretical theorems and treatment frameworks of UGC/UGD [2,64] and CATA [47,48]. Specifically, we used Krippendorff's [75,80] assessments for the content analysis of the sample. In what follows, each of the developed procedures and the extraction of the sample are explained in further detail.

3.1. Data sampling

As outlined above, in the present study, a three-phase methodological approach was used. First, we performed sentiment analysis that works with machine learning to subdivide a database of 938,258 Tweets containing information related to tweets published under the UGC framework containing the hashtags #InternetofThings, #HomeAssistant or #IoT.

The tweets were collected from the Twitter API from November 1 to December 11, 2020. At the time of data collection, there were no events in the sector that could differ in the industry related to the Internet of Things and HomeAssistant, such as big fairs or events [81]. The results were also filtered based on the elimination of repeated tweets and retweets. Only the texts appearing in the collected tweets were analyzed, as the present study is based on the Natural Language Processing (NLP) framework that focuses on text analysis as a source of information [82]. Python and Pandas libraries were used for tweet extraction and filtering (see [83]).

In this way, a total of 80,466 repeated tweets and 111,334 retweets are obtained. Therefore, the final sample after the database debugging and filtering process amounted to 746,458 tweets containing the hashtags #InternetofThings, #HomeAssistant or #IoT. The database of 746,458 tweets was used to develop the experimental process of CATA [48] based on a UGC sample [84].

3.2. Sentiment analysis

In sentiment analysis, different approaches are used to subdivide the data into categories that express different emotions. In the present study, following the standard procedure in UGC and CATA analysis, we relied on the NLP framework [82] and focused exclusively on text,

Table 1
Relevant studies.
Source: The authors.

Authors	Aims
Alamsyah et al. [51]	This paper argued that Twitter can serve as a source of real-time thoughts and opinions from users that can be used to map the public opinion towards a topic.
Mishra et al. [66]	This study tracked different opinions about a product to analyze them and classify them into positive, negative, or neutral sentiments.
Liu et al. [67]	This study unidentified a model to cope with noisy labels, such as emoticons for a clear behavioral classification.
Janssen et al. [68]	In this study, key IoT challenges related to security, privacy, and data quality were recognized.
Komninos et al. [69]	This paper categorized threats to the security smart grid and home environment to provide countermeasures to preserve user security.
Pecorella et al. [70]	In this study, we authors developed a network sentiment analysis to dynamically adapt the security level of the smart home network.

thus disregarding emoticons, images, URLs, or any other multimedia elements contained in the tweets.

Textblob was used for sentiment analysis. This procedure is commonly used in academic research [85]. Textblob is a well-known library developed in Python programming language and is used to find common text processing operations [77]. Textblob is built on NLTK and patterns [86]. While one of the limitations of sentiment analysis is that it is challenging to apply it to texts that contain connotations, sarcasm, irony, and so forth, in the present study, we took issues into account and tried to proceed with the development of the methodology so that to avoid these challenges [87].

Therefore, the results showed a polarity score that was measured and classified as polarity or subjectivity. Polarity ranged from −1 to 1, and subjectivity from 0.0 to 1.0 [85]. The algorithm that works with sentiment analysis was trained a total of 744 times using tweets that were manually classified, giving the algorithm in-puts centered on sentiment classification [88]. The algorithm, based on these questions, learned by itself, and the more tweets it analyzed, the higher was its success rate, since it works with machine learning [84].

Once tweets were divided into different sentiments, we obtained three databases of tweets expressing negative, positive, and neutral sentiments. Although sentiment analysis algorithms can indeed divide the databases into a multitude of sentiments, in the present study, we decided to use only three sentiments, because our study objective could be met by using this standard in the NLP approach methodology.

The article was validated for 5 cross-validations as indicated in Hiremath and Patil [85]. For the measurement of the results, the factors of precision, recall, f1-score, and support were considered (see Section 4). The results in terms of macro average and weighted average were also taken into account. For the classification of patterns in the analyzed datasets, we used the following category models:

- Support vector classifier [89]
- Multinomial naïve Bayes [90]
- Logistic regression [89]
- Random forest classifier [91]

3.3. Latent Dirichlet Allocation (LDA)

In the next step, the three databases were analyzed with a topic-modeling algorithm known as LDA [92]. LDA enables the analysis of the words that compose the analyzed documents. The LDA algorithm was developed in Python LDA 1.0.5 using Gibbs sampling (MAC version). Using the LDA algorithm, we obtained major themes in each of the three datasets [78].

These themes were consequently divided into feelings, and theme names were manually composed of frequently occurring words in each of the themes [93]. This is a standard approach used in previous research on modeling themes based on text analysis [94].

Overall, the thematic modeling algorithm aims to study the different inputs understood as the words contained in the different databases based on frequency and positioning of words in the documents (Parra and Santander, 2015). In this way, a percentage of the relevance of each of these words was computed and based on these values, we identified and names the themes in the data [95]. As argued by Khan et al. [96], an LDA model is a probabilistic assumption developed in two parts. In the first part, words were identified in separate documents (i.e., sentiment databases).

Next, two consecutive steps were performed: in the first step, a distribution of themes in the sample was identified; in the second step, the themes were automatically grouped into keyword forms according to the number of times they were repeated. The equation applied with Python is shown in Eq. (1).

$$\rho\left(\beta_{1:k},\theta_{1:D},Z_{1:D},\omega_{1:D}\right)=\prod_{i=1}^{K}\rho(\beta_i)\left(\beta_1\right)\times\prod_{d=1}^{D}\rho\left(\theta_d\right)\times\sum_{n=1}^{N}\rho(Z_{d,n}|\,\theta_d)\rho\left(W_{d,n}|\,\beta 1_{:K},Z_{d,n}\right) \quad (1)$$

β_i *Distribution of word in topic i, altogether K topics*
θ_d *Proportions of topics in document d, altogether D documents*
z_d *Topic assignment in document d*
$z_{d,n}$ *Topic assignment for the nth word in document d, altogether N words*
w_d *Observed words for document d*
$w_{d,n}$ *The nth word for document d*

Next, the automatic division of the words was applied to the analysis of the themes using the approximation shown in Eq. (2). Based on the results, we analyzed the first 10 words of each sub-theme to create the names of the themes [84].

$$\rho\left(\beta_{1:k},\theta_{1:D},Z_{1:D}|\,\omega_{1:D}\right)=\frac{\rho\left(\beta_{1:K},\theta_{1:D},z_{1:D},\omega_{1:D}\right)}{p\left(w_{1:D}\right)} \quad (2)$$

3.4. Textual analysis

Finally, we applied textual analysis to calculate the weighted percentages of a keyword in the analyzed dataset. The textual analysis approach takes into consideration the analysis of the weight of the words in the entire database. In this way, a percentage of the relevance of those words in the database can be evaluated [97]. Based on this percentage, researchers can extract insights about the patterns identified in the database or extract considerations about the object of study to analyze them from a specific perspective [98].

In the present study, indicators such as weight percentage and the number of times a word was repeated were considered. This exploratory approach enabled extracting insights to address the research questions a [84].

Furthermore, using a well-known approach in NLP [99], we identified additional n-grams collected from the three datasets. Using Latin numerical prefixes, an n-gram of size 1 was referred to as a "unigram"; size 2 was a "bigram". Therefore, n-grams relative to unigram (positive, negative, and neutral) as well as bigram (negative, positive, and neutral) were identified in the process of textual analysis. In doing so, we followed Reyes-Menendez et al. [100].

In this way, in other to compute the n-grams analysis we followed Wu and Su [101] whose validated their hypothesis using the statistical measure known as mutual information (MI) for n-grams. According to Reyes-Menendez et al. [100] this statistical measure refers to the probability of co-occurrence of two indicator variables that are perfectly correlated.

Table 2
Model category details.

Sl. No.	Model Name	Fold_idx	Accuracy - Textblob
0	RandomForestClassifier	0	0.511456
1	RandomForestClassifier	1	0.524034
2	RandomForestClassifier	2	0.530548
3	RandomForestClassifier	3	0.542677
4	RandomForestClassifier	**4**	**0.555481**
5	LinearSVC	0	0.862758
6	LinearSVC	1	0.859838
7	LinearSVC	2	0.865229
8	LinearSVC	**3**	**0.869946**
9	LinearSVC	4	0.866352
10	Multinomial Naïve Bayes	0	0.719677
11	Multinomial Naïve Bayes	1	0.701482
12	Multinomial Naïve Bayes	**2**	**0.739892**
13	Multinomial Naïve Bayes	3	0.736523
14	Multinomial Naïve Bayes	4	0.730683
15	LogisticRegression	0	0.836703
16	LogisticRegression	1	0.827942
17	LogisticRegression	**2**	**0.839173**
18	LogisticRegression	3	0.836029
19	LogisticRegression	4	0.836029

Authors such as Bouma [102] and Iyengar et al. [103] used MI to validate whether there is a shared correlation between words and this measurement must be compared to the indicator frequency (F). Following MI value between random variables X and Y, whose values have marginal probabilities as stated by Reyes-Menendez et al. [100], and p(x) and p (y), and joint probabilities p(x, y), can be computed using Eq. (3).

$$I(X;Y) = \sum_{x,y} p(x,y) \mathrm{In} \frac{\rho(x,y)}{p(x)\,p(y)} \tag{3}$$

A summary of the methodological process is shown in Fig. 1 which shows the steps of data collection and the methodological process. Regarding the sample, Fig. 1 shows the data collection process in Twitter and its collection date. The use of SA and Textblob was tested with SVC, MNB, LG, and RFC. Next, topic modeling with LDA was used to obtain the sample's sentiments, which were then validated with keyness, *p*-value, and the topic's identification (including the name, sentiment, and descriptions). Next, we performed of TA using the NLP perspective and identified n-grams. In this step, WP metrics, total keywords count, and the n-gram's sentiments were calculated. Finally, the results were discussed based on the UGD and UGC theoretical perspective; and the considerations of the content analysis were proposed in the CATA conceptual framework. Fig. 1 also shows additional information related to the use of Python and Textblob.

4. Results

4.1. Results of sentiment analysis

The experiment and approximation were modeled with the support of the standard classifier methods namely, logistic regression, naïve Bayes, SVC, and random forest classifier. Accuracy measured the percentage of cases when model succeeded. This is one of the most used metrics in machine learning studies [99]. The highest accuracy result in this study was found about Linear SVC Sl. No. 8 (0.869946) and 9 (0.866352). As concerns random forest classifier, the highest accuracy was 0.555481.

The accuracy of multinomial Naïve Bayes amounted to 0.739892, while that of logistic regression was 0.839173, being this last value of accuracy the higher found in the analysis.

Table 2 shows the results of the classification experiments, highlighting the highest accuracy values achieved using Textblob in each of the models.

Table 3 summarizes brief scores of the Textblob analysis according to the model used. IAs can be seen in Table 3, the highest values to

Table 3
Summarized brief scores.

Sl. No.	Model Name	Scores of Textblob analysis
1	LinearSVC	**0.864825**
2	LogisticRegression	**0.835175**
3	MultinominalNB	0.725651
4	RandomForestClassifier	0.532839

Table 4
Classification report.

Sl. No.	Parameters	Vader			
		precision	recall	f1-score	support
1	Negative	0.74	**0.81**	0.74	20.511
2	Positive	0.84	0.75	0.79	2301
3	Neutral	0.89	**0.93**	0.92	20.493
4	Accuracy	–	–	0.85	43.642
5	Macro avg	0.79	0.75	0.73	43.642
6	Weighted avg	0.79	0.82	0.84	43.642

the set of accuracy in the results are those corresponding to linear SVC and logistic regression, with the values 0.864825 and 0.835175, respectively.

The classification report presented in Table 4 shows the positive, negative, or neutral accuracy by Textblob, as well as the recall value and the fl-score and support values. Accuracy is a metric that measures the quality of the machine learning model in classification tasks. Recall measures reflect the quantity that the machine learning model can identify in a database. F1-score is used to combine precision and recall measurements into a single value, which is a practical approach, because it facilitates comparing the combined performance of accuracy and completeness between various solutions. Finally, support measures reflect the support of the machine learning machine used to predict the model. Macro average measures the total average of the model based on the variables analyzed, while weighted average measures its relativity in terms of weight. The highest recall values were obtained for negative and neutral tweets, with the corresponding values of 0.81 and 0.91, respectively. Positive tweets obtained a 0.75 recall.

4.2. LDA results

Using the LDA algorithm, we identified a total of 10 topics related to the security or privacy of IoT and home assistants in smart living environments. For each of the topics identified in the databases subdivided into feelings, the first 20 topics of each database were analyzed. To this end, the first 9–12 words that appeared by relevance as a result of the automated classification process after applying the LDA were identified.

The topics were named based on the words grouped within them. In the naming process, we tried to use the most frequent words to form a title, which is standard approach in LDA studies [104].

Similarly, the topics from which the insights were extracted based on the sentiment expressed in tweets were named based on the study of the terms and concepts used in the scientific literature on privacy and security of IoT devices in smart living environments. From the total number of topics identified in the LDA process, the topics indicated in Table 6 were manually selected to answer the research questions.

Furthermore, in order to measure the relevance of the identified topics, their keyness values were calculated. The keyness is defined as the strength of the link between the topics; statistically, this measure determines the log-likelihood score values [105].

In the present study, the keyness was used to measure the statistical relevance of the topics in the complete database of the tweets upon filtering. Of note, log-likelihood of >3.8 was statistically significant when p-value < 0.05. Table 5 summarizes the results regarding the topics, their descriptions, associated sentiment, keyness, and p-values.

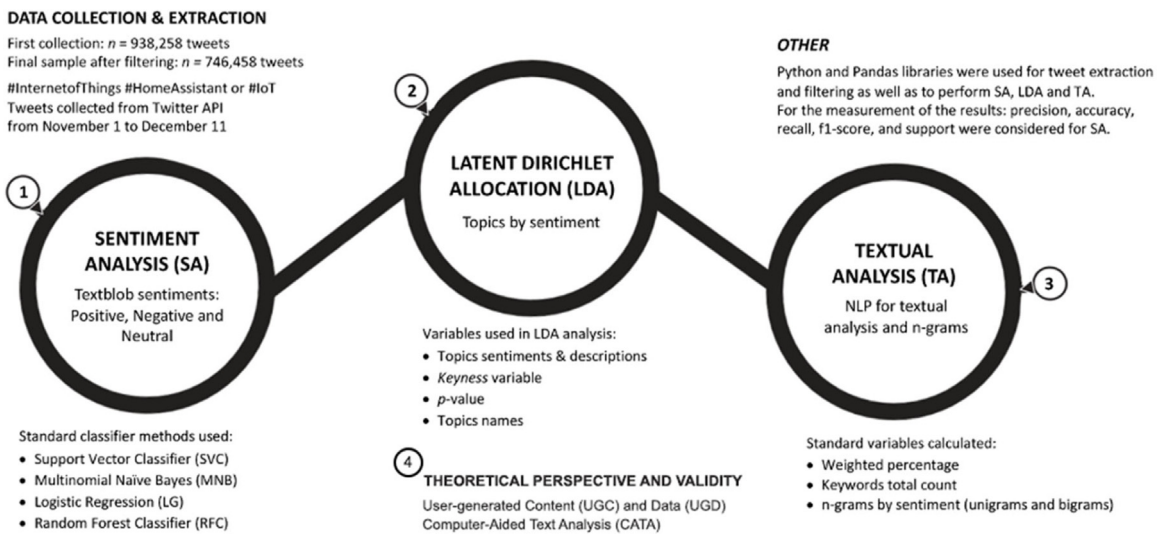

Fig. 1. Summary of the methodological process.

Table 5
LDA topics results sentiment, Keynes and *p*-value.

R	Topics	Description	Sent.[a]	Keyness	*p*-value
1	IoT devices	Summary of a multitude of content about IoT devices and connected devices	N	803.05	0.041
2	Malware	Different malware processes found in IoT devices	Ne	692.90	0.039
3	Cybersecurity attacks	Summary of topics related to the analysis of cyber-attacks	Ne	601.23	0.037
4	Cloud-based security	Cloud-based security systems to which the devices are connected in smart home environments	P	591.83	0.032
5	Storing of data	Access to databases that collect user information	Ne	583.04	0.031
6	Testing software	It Systems in test or users who install smart devices in their homes	Ne	467.70	0.028
7	Data Leaks	Companies misusing user data that may put users at risk	Ne	461.15	0.027
8	Anonymize data	Initiatives to support anonymous user data and anonymization processes	P	397.84	0.025
9	Cloud companies	Companies that store information concerning IoT devices in the cloud	N	392.75	0.024
10	Medical technology	Digital healthcare and telemedicine-related technologies	N	145.21	0.019

[a]Sent. = Sentiment, N = Neutral sentiment; Ne = Negative sentiment; P = Positive sentiment.

Table 6
Grouped keywords, count, and weighted percentage.

R	Word	Similar words	Frequency	WP
1	IoT	connected devices, IoT devices for home, programmable IoT devices, IoT sensors, etc.	130772	18.11
2	Malware	Malware, malicious software, IoT malware, IoT ransomware threat, etc.	23286	9.78
3	Cybersecurity	Cybersecurity attacks, IoT devices attacks, trending IoT malware attack, protect IoT devices, etc.	19019	6.59
4	Cloud	Cloud-based security, cloud computing, secure integration, cloud-based IoT security, etc.	17024	5.21
5	Data store	Storing of data, IoT data storage, IoT data deluge, data store issues, router data store, etc.	17009	5.21
6	Test	Testing software, IoT testing software, IoT testing tools, IoT testing, wireless IoT testing, etc.	16731	4.89
7	Data leaks	Data leaks, IoT private data, IoT data breaches, security breaches, IoT-related data leaks, etc.	16240	4.16
8	Anonymize	Anonymize IoT data, IoT data streams, IoT-based anonymous functions, anonymization in IoT, etc.	15801	3.67
9	Cloud companies	IoT cloud computing, cloud platforms, management of IoT devices, etc.	14010	3.09
10	e-Health	Medical technology, digital health services, IoT medical care, e-health, IoT medical devices, etc.	13091	2.99

4.3. Results of textual analysis

In order to proceed with the textual analysis process, we analyzed the number of times that the keywords were repeated in the databases and the weighed percentage in the total database [97]. To this end, we used the NLP framework using Pandas GroupBy in Python.

In this way, the count variables relative to the number of times the words were repeated and their frequency in the database were obtained. Table 6 shows similar keyword sets and the total weight in the database.

Next, we analyzed the n-grams in the databases as one of the procedures highlighted in CATA-based studies. An n-gram model predicts the occurrence of a word based on the occurrence of its $n-1$ previous word. Similarly, the bigram model ($n = 2$) predicts the occurrence of a word given only its previous word (as $n - 1 = 1$). Based on this approach, Table 7 lists n-grams divided into feelings for unigrams and bigrams. Moreover, the analyzed n-grams are supported in placement, allowing researchers to take into account the contexts where words occur in a corpus [106,107]. The context is defined as words that are usually used together. Reyes-Menendez et al. [100] indicated that, if placement presents a strong and stable relationship, it is called a lexical or n-gram package. In Tables 7 to 16, each of the topics is presented per Rank (R) with the keywords identified in Table 6.

We also list the terms that usually accompany these terms according to the objectives of the study and therefore removing from the results those terms that did not fit in to the study research questions as were categorized as not inclusive. Likewise, in this case the term Freq/count refers to the total frequency of appearance of the collocates. This is the sum of FreqL of words that appear on the left on the topic and FreqR of words that appear on the right of the topic [100].

5. Discussion

The recent massive use of IoT devices, home assistants, and new intelligent systems has led to the development of new privacy concerns,

Table 7
N-grams for the collocates of the neutral topic "IoT".

R	Collocates for IoT			
	Freq	Freq L	Freq R	Collocate
1	13336	7901	5435	#IoT
2	7912	134	104	ConnectedDevices
3	2012	780	1232	IoTsensors
4	1003	450	553	Programmable
5	750	359	391	Services

Table 8
N-grams for the collocates of the negative topic "Malware".

R	Collocates for Malware			
	Freq	Freq L	Freq R	Collocate
1	6840	2905	3935	#IoT
2	2081	1032	1978	Malware
3	1145	573	572	Threats
4	791	379	412	Malicious
5	234	130	104	Attacks

Table 9
N-grams for the collocates of the negative topic "Cybersecurity".

R	Collocates for Cybersecurity			
	Freq	Freq L	Freq R	Collocate
1	4091	2074	2017	#IoT
2	3301	1505	1796	IoTattacks
3	2781	1407	1374	IoTsecurity
4	491	249	242	Protections
5	401	141	260	Cyberattacks

Table 10
N-grams for the collocates of the positive topic "Cloud".

R	Collocates for Cloud			
	Freq	Freq L	Freq R	Collocate
1	3911	2842	1069	#IoT
2	2415	1905	510	CloudIoT
3	1703	790	913	Cloudcomputing
4	804	394	410	Cloudsecurity
5	200	106	94	Integrations

Table 11
N-grams for the collocates of the negative topic "Data store".

R	Collocates for Data store			
	Freq	Freq L	Freq R	Collocate
1	3701	1890	1811	#IoT
2	3501	1703	1798	Datastoreage
3	2756	1807	949	Attacks
4	1429	559	870	Issues
5	260	115	145	Router

Table 12
N-grams for the collocates of the negative topic "Test software".

R	Collocates for Test software			
	Freq	Freq L	Freq R	Collocate
1	2591	1590	1001	#IoT
2	2390	1361	1230	Tests
3	301	194	107	Testingtools
4	297	174	123	Wireless
5	241	189	52	Sensors

Table 13
N-grams for the collocates of the negative topic "Data leaks".

R	Collocates for Data leaks			
	Freq	Freq L	Freq R	Collocate
1	2101	1048	1997	#IoT
2	1801	953	848	Breaches
3	1791	874	917	Leaks
4	1090	671	419	Privacy
5	749	344	405	Users

Table 14
N-grams for the collocates of the positive topic "Anonymize".

R	Collocates for Anonymize			
	Freq	Freq L	Freq R	Collocate
1	1994	793	1201	#IoT
2	1880	781	1099	Anonymize
3	1803	808	995	Data
4	489	389	100	Functions
5	371	156	215	Privacy

Table 15
N-grams for the collocates of the neutral topic "Cloud companies".

R	Collocates for Cloud companies			
	Freq	Freq L	Freq R	Collocate
1	1031	501	530	#IoT
2	1001	701	300	Cloud
3	957	646	311	Computing
4	401	276	125	Management
5	230	129	101	Platforms

Table 16
N-grams for the collocates of the neutral topic "e-Health".

R	Collocates for e-Health			
	Freq	Freq L	Freq R	Collocate
1	976	400	576	#IoT
2	890	549	341	Medical
3	807	401	406	Health
4	602	309	293	Services
5	501	270	231	Hardware

both for users and companies that manage user data [108]. As argued by Kong et al. [109], the currency of the 21st century is data and the insights that can be extracted from data analysis with the techniques focused on data automation, machine learning, or artificial intelligence.

In line with similar findings reported by Kumar and Patel [110] and Mohammad [111], the results of the present study revealed that there are concerns regarding how data obtained from connected devices are managed. These concerns are the main neutral topic identified in this study (keyness = 803.05; p = 0.041), with the textual analysis frequency of 130.772 and a WP of 18.11. This topic (IoT) stood out in our analysis of n-grams of the collocates the features of programmable (Freq 1003) IoT devices and services (Freq 750).

However, with regard to the proposed research objectives, we found security issues related to malware (keyness = 692.90; p = 0.039) that can be used in connected devices in smart living environments. The textual analysis results revealed the importance of malicious software, a specific malware for home assistants, or specific attacks such as ransomware threat (total frequency 23286 keywords and a WP of 9.78). These concerns were also identified by Arabo et al. [112]. Also, n-grams for the collocates of the negative topic malware highlight the specific concerns about malware (Freq 2081), threats (Freq 1145), malicious (Freq 791), and attacks (Freq 234).

Similarly, the results of topic-based approach suggest the concerns related to the issues regarding cyber-attacks (keyness = 601.23; p = 0.0379) that could target connected devices in smart living environments. These issues highlight the relevance of user personal data when they use these IoT devices in their homes [113], as suggested by the results of n-grams for the collocates of negative topics related to data leaks, such as breaches (Freq 1801), leaks (Freq 1791), and privacy (Freq 1090).

Furthermore, since the IoT device data are usually located and stored in cloud storage systems [114], these cloud systems can also

be attacked, and user personal data can reach cybercriminals [115]. In our results, this topic was associated with positive sentiment (keyness = 591.83, $p = 0.032$), in contrast to the negative topic related to standard storing of data (keyness = 583.04, p-value = 0.031. As argued by Patel and Doshi [116] cloud-centric systems – regardless of whether or not the data come from IoT – are usually one of the points that receive the most attacks from cyber-criminals, as such data can be used for extortion or identity theft [117]. Databases that store user data are a relevant source of data not only for companies that manage and analyze these data, but also for users and their confidence and trust in using the Internet products [118]. In our results, the topic of cloud storage systems had the frequency of 17024 and WP of 5.21.

Furthermore, we also observed the relevance of the data storage capacity from companies that work with cloud services linked to IoT devices, with the identification of the neutral topic cloud companies (keyness = 392.75; $p = 0.024$). Therefore, data management by companies is a priority for the user data security. As suggested by the n-grams of neutral topic cloud companies, the data management (Freq 401) and the use of digital platforms (Freq 230) for data filtering and storage are relevant to users.

Furthermore, in our results, we also identified a negative topic that focuses on the use of IoT in the test mode (keyness = 467.70; $p = 0.028$). Test modes are connected products launched on the market that, owing to machine learning technology, improve over time [119]. Sometimes, these devices can cause problems in their operation, although users are aware that they are using an emerging technology that will improve and get more intelligent over time [120]. The textual analysis results showed relative occurrences to test modes of software, hardware, wireless among other IoT devices (Freq 16731 keywords, WP of 4.89). In n-grams for the negatives collocates, testing tools (Freq 301), wireless (Freq 297) and sensors (Freq 241) gained relevance.

Another concern we identified in the results is that of data leaks (see also [121]). Both companies and users are aware that their accounts might be hacked, both on the personal level through their accounts and through the routers in the smart living environments where the devices are installed [122]. Therefore, security concerns related to these data leaks are relevant for the industry (keyness = 461.15, $p = 0.027$) [113]. This negative identified topic links security concerns regarding anonymity and data. The data anonymity has been identified as a positive topic (keyness = 392.75; $p = 0.024$) where, in the study of the n-grams, the functionalities (Freq 489) and the increase of privacy (Freq 372) stood out. However, while this is a regular process undertaken by companies, there are concerns about whether user personal data remain anonymous as was studied by Madaan et al. [123] and Ishmaev [124]. Our results showed the relevance of these concerns (data leaks topic), with the frequency of 16240 keywords and WP of 4.16. Moreover, we found a direct relationship between the keywords "breaches" (Freq 1801) and "leaks" (Freq 1791) and between "privacy" (Freq 1090) and "users" (Freq 749).

In addition, an important insight offered by our results is related to the use of IoT and home assistant systems in digital health services. As suggested by the identified neutral topic "medical technology", these technologies are being used as novel forms of support in hardware and software technologies to generate data sources However, this topic was the least relevant in our data (keyness = 145.21, $p = 0.019$).

Furthermore, the IoT allows for the development of new functionalities and features that have not been previously used in the health or insurance industry [125]. The results of our textual analysis highlighted the importance of digital health services, medical care, or the use of the e-health app (Freq 13091; WP 2.99), which was previously studied previously by Khemissa and Tandjaoui [126] and Vilela et al. [127]. Likewise, the relevance of this topic to the smart living industry was confirmed by n-grams for the collocates—namely, medical (Freq 890), health (807), services (602) and hardware (501).

Accordingly, the strategies focused on the understanding of identity management should be prioritized for further development in the short-term future. When buying IoT devices for smart living environments, users should understand the technology and risks of buying these connected products. These initiatives should be carefully considered by companies that develop these devices and public agencies that work to educate users on the risks of transforming their homes into smart living environments.

6. Conclusions

In the present study, we used a three-phase methodology to extract insights regarding security issues of IoT devices in smart living environments using data-mining techniques. Based on the results, we identified 10 topics and classified the analyzed tweets into those expressive three different types of sentiment (positive, negative, and neutral). More specifically, the positive topics were cloud-based security and anonymize data; furthermore, negative topics included malware, cybersecurity attacks, storing of data, testing software and data Leaks; finally, neutral topics were IoT devices, cloud companies, and medical technology. We also extracted insights regarding the main keywords that composed the topics and explored the most correlated n-grams.

Based on the results, with regard to our first research question (*What are the main security issues of IoT devices in smart living environments?*), we found that the main security issues are malware, cybersecurity attacks, data storing vulnerabilities, the use of testing software in IoT, and possible leaks due to the lack of user experience.

Furthermore, regarding RQ2 (*How can security issues of IoT devices in smart living environments affect the privacy of users who use these technologies?*), using the insights from textual analysis and n-grams, as well as the identified topics, we discussed different circumstances and causes that may affect user security and privacy when using IoT devices. Our results also emphasize the importance of UGC in the processing of personal data of IoT device users, access to medical information or disease care, personal habits in home environments or illegal access to unwanted information by users of these devices has been highlighted.

Finally, our results also revealed the importance of identity management in IoT devices used in smart living environments. In this way, users can be aware of the data they generate as well as of potential vulnerabilities of the systems, possible gaps, or information leaks.

Theoretical implications

Further research can use the topics identified in the present study to establish constructs, variables, or indicators that specifically study IoT privacy risk issues, both within and beyond smart living environments. In addition, the present study proposes a novel methodology that contributes to the IoT sector. Therefore, based on the methodological development focused on data mining and the use of the UGC as a data source, researchers can develop similar approaches to find insights and patterns in relation to security issues using the content shared by users on the Internet and social networks.

Similarly, the insights through textual analysis and the results of our study of n-grams enable creating substantiate research proposals and hypotheses to be tested using similar approaches or quantitative methods to establish their conclusions regarding the IoT industry in future research.

Practical implications

With regard to practical implications of our findings, agencies and practitioners can use the results of the present study to establish new security protocols not only at the technical level, but also at the level of educational communication, so that to establish fluid channels of information with users who buy their products and services.

Users should be aware of the security issues that can result from failed installations of these devices or their misuse. In addition, companies can use our results to better understand users' appreciation of the security of IoT connected devices, improve their products, or make users thoroughly understand the risks of excessive use of such devices.

Limitations and future research

The limitations of the present study are related to the size of the sample and the time horizon analyzed. Another possible limitation of this work is our use of exploratory approaches that work with machine

learning. The more these approaches are trained, the better their predictive ability. The LDA process using which the identified topics were named could also a limitation of the present study, since it is a manual process developed within the standard LDA model approach.

In further research, it would be necessary to focus on improving the analysis processes used in this study, as well as identify new issues related to safety of IoT devices and home assistants and user privacy in smart living environments.

CRediT authorship contribution statement

Jose Ramon Saura: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Daniel Palacios-Marqués:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Domingo Ribeiro-Soriano:** Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

In gratitude to the Ministry of Science, Innovation and Universities, Spain and the European Regional Development Fund: RTI2018-096295-B-C22.

References

[1] N. Seydoux, K. Drira, N. Hernandez, T. Monteil, IoT-O, a core-domain IoT ontology to represent connected devices networks, in: European Knowledge Acquisition Workshop, Springer, Cham, 2016, pp. 561–576, http://dx.doi.org/10.1007/978-3-319-49004-5_36.

[2] S. Ribeiro-Navarrete, J.R. Saura, D. Palacios-Marqués, Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy, Technol. Forecast. Soc. Change 167 (2021) 120681, http://dx.doi.org/10.1016/j.techfore.2021.120681.

[3] S. Kraus, N. Roig-Tierno, R.B. Bouncken, Digital innovation and venturing: An introduction into the digitalization of entrepreneurship, Rev. Manag. Sci. 13 (3) (2019) 519–528.

[4] S.R. Sahoo, B.B. Gupta, Multiple features based approach for automatic fake news detection on social networks using deep learning, Appl. Soft Comput. 100 (2021) 106983, http://dx.doi.org/10.1016/j.asoc.2020.106983.

[5] V. Sivaraman, H.H. Gharakheili, C. Fernandes, N. Clark, T. Karliychuk, Smart IoT devices in the home: Security and privacy implications, IEEE Technol. Soc. Mag. 37 (2) (2018) 71–79, http://dx.doi.org/10.1109/MTS.2018.2826079.

[6] R. Bouncken, Y. Qiu, The impact of digitalization on organizations - A review of the empirical literature, Int. J. Entrepreneurial Ventur. 2021 (2021) http://dx.doi.org/10.5465/AMBPP.2019.19324abstract.

[7] F. Kawsar, A.B. Brush, Home computing unplugged: why, where and when people use different connected devices at home, in: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, 2013, September, pp. 627–636.

[8] T. Alladi, V. Chamola, B. Sikdar, K.K.R. Choo, Consumer IoT: Security vulnerability case studies and solutions, IEEE Consum. Electron. Mag. 9 (2) (2020) 17–25, http://dx.doi.org/10.1109/MCE.2019.2953740.

[9] N. Seydoux, K. Drira, N. Hernandez, T. Monteil, IoT-O, a core-domain IoT ontology to represent connected devices networks, in: European Knowledge Acquisition Workshop, Springer, Cham, 2016, pp. 561–576, http://dx.doi.org/10.1007/978-3-319-49004-5_36.

[10] R. Bouncken, R. Barwinski, Shared digital identity and rich knowledge ties in global 3D printing - A drizzle in the clouds? Glob. Strategy J. (2020) 1–28, http://dx.doi.org/10.1002/gsj.1370.

[11] J.P. Nzabahimana, Analysis of security and privacy challenges in Internet of Things, in: 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies, DESSERT, IEEE, 2018, pp. 175–178, http://dx.doi.org/10.1109/DESSERT.2018.8409122.

[12] J.R. Saura, Using data sciences in digital marketing: Framework, methods, and performance metrics, J. Innov. Knowl. 6 (2) (2021) 92–102, http://dx.doi.org/10.1016/j.jik.2020.08.001, April–June 2021.

[13] C.M. Medaglia, A. Serbanati, An overview of privacy and security issues in the internet of things, Internet Things 38 (2010) 9–395, http://dx.doi.org/10.1007/978-1-4419-1674-7_38.

[14] N. Zainuddin, M. Daud, S. Ahmad, M. Maslizan, S.A.L. Abdullah, A study on privacy issues in internet of things (IoT), in: 2021 IEEE 5th International Conference on Cryptography, Security and Privacy, CSP, IEEE, 2021, pp. 96–100.

[15] L.A. Tawalbeh, F. Muheidat, M. Tawalbeh, M. Quwaider, IoT privacy and security: Challenges and solutions, Appl. Sci. 10 (12) (2020) 4102, http://dx.doi.org/10.3390/app10124102.

[16] C. Esposito, M. Ficco, B.B. Gupta, Blockchain-based authentication and authorization for smart city applications, Inf. Process. Manage. 58 (2) (2021) 102468, http://dx.doi.org/10.1016/j.ipm.2020.102468.

[17] C.D. Nugent, X. Hong, J. Hallberg, D. Finlay, K. Synnes, Assessing the impact of individual sensor reliability within smart living environments, in: 2008 IEEE International Conference on Automation Science and Engineering, IEEE, 2008, pp. 685–690, http://dx.doi.org/10.1109/COASE.2008.4626548.

[18] J.M. Blythe, N. Sombatruang, S.D. Johnson, What security features and crime prevention advice is communicated in consumer IoT device manuals and support pages? J. Cybersecur. 5 (1) (2019) http://dx.doi.org/10.1093/cybsec/tyz005, tyz005.

[19] M. Ullah, P.H. Nardelli, A. Wolff, K. Smolander, Twenty-one key factors to choose an iot platform: Theoretical framework and its applications, IEEE Internet Things J. 7 (10) (2020) 10111–10119, http://dx.doi.org/10.1109/JIOT.2020.3000056.

[20] Z.K. Zhang, M.C.Y. Cho, C.W. Wang, C.W. Hsu, C.K. Chen, S. Shieh, IoT security: ongoing challenges and research opportunities, in: 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications, IEEE, 2014, pp. 230–234.

[21] S. Ammirato, F. Sofo, A.M. Felicetti, C. Raso, A methodology to support the adoption of IoT innovation and its application to the Italian bank branch security context, Eur. J. Innov. Manag. (2019) http://dx.doi.org/10.1108/EJIM-03-2018-0058.

[22] K.K. Patel, S.M. Patel, Internet of things-IOT: definition, characteristics, architecture, enabling technologies, application & future challenges, Int. J. Eng. Sci. Comput. 6 (5) (2016).

[23] R. Bouncken, S. Kraus, Entrepreneurial ecosystems in an interconnected world: Emergence, governance, and digitalization, Rev. Manag. Sci. (2021) http://dx.doi.org/10.1007/s11846-021-00444-1.

[24] A. Hosseinian-Far, M. Ramachandran, C.L. Slack, Emerging trends in cloud computing, big data, fog computing, IoT and smart living, in: Technology for Smart Futures, Springer, Cham, 2018, pp. 29–40, http://dx.doi.org/10.1007/978-3-319-60137-3_2.

[25] H. Lin, N.W. Bergmann, IoT privacy and security challenges for smart home environments, Information 7 (3) (2016) 44, http://dx.doi.org/10.3390/info7030044.

[26] R. Chow, The last mile for IoT privacy, IEEE Secur. Priv. 15 (6) (2017) 73–76, http://dx.doi.org/10.1109/MSP.2017.4251118.

[27] S. Sicari, A. Rizzardi, L.A. Grieco, A. Coen-Porisini, Security, privacy and trust in Internet of Things: The road ahead, Comput. Netw. 76 (2015) 146–164, http://dx.doi.org/10.1016/j.comnet.2014.11.008.

[28] X. Caron, R. Bosua, S.B. Maynard, A. Ahmad, The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective, Comput. Law Secur. Rev. 32 (1) (2016) 4–15, http://dx.doi.org/10.1016/j.clsr.2015.12.001.

[29] P.P. Jayaraman, X. Yang, A. Yavari, D. Georgakopoulos, X. Yi, Privacy preserving Internet of Things: From privacy techniques to a blueprint architecture and efficient implementation, Future Gener. Comput. Syst. 76 (2017) 540–549, http://dx.doi.org/10.1016/j.future.2017.03.001.

[30] M. Ma, P. Wang, C.H. Chu, Data management for internet of things: Challenges, approaches and opportunities, in: 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, IEEE, 2013, pp. 1144–1151.

[31] S. Gupta, B.B. Gupta, XSS-secure as a service for the platforms of online social network-based multimedia web applications in cloud, Multimedia Tools Appl. 77 (4) (2018) 4829–4861, http://dx.doi.org/10.1007/s11042-016-3735-1.

[32] M.U. Hassan, M.H. Rehmani, J. Chen, Privacy preservation in blockchain based IoT systems: Integration issues, prospects, challenges, and future research directions, Future Gener. Comput. Syst. 97 (2019) 512–529, http://dx.doi.org/10.1016/j.future.2019.02.060.

[33] Y.N. Liu, Y.P. Wang, X.F. Wang, Z. Xia, J.F. Xu, Privacy-preserving raw data collection without a trusted authority for IoT, Comput. Netw. 148 (2019) 340–348, http://dx.doi.org/10.1016/j.comnet.2018.11.028.

[34] S.S. Gill, P. Garraghan, R. Buyya, ROUTER: Fog enabled cloud based intelligent resource management approach for smart home IoT devices, J. Syst. Softw. 154 (2019) 125–138, http://dx.doi.org/10.1016/j.jss.2019.04.058.

[35] B.B. Gupta, M. Quamara, An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols, Concurr. Comput.: Pract. Exper. 32 (21) (2020) e4946, http://dx.doi.org/10.1002/cpe.4946.

[36] A. Chaudhuri, Internet of things data protection and privacy in the era of the General Data Protection Regulation, J. Data Protect. Priv. 1 (1) (2016) 64–75.
[37] H. Tao, M.Z.A. Bhuiyan, M.A. Rahman, G. Wang, T. Wang, M.M. Ahmed, J. Li, Economic perspective analysis of protecting big data security and privacy, Future Gener. Comput. Syst. 98 (2019) 660–671, http://dx.doi.org/10.1016/j.future.2019.03.042.
[38] J.H. Kim, A survey of IoT security: Risks, requirements, trends, and key technologies, J. Ind. Integr. Manag. 2 (02) (2017) 1750008, http://dx.doi.org/10.1142/S2424862217500087.
[39] F.K. Santoso, N.C. Vun, Securing IoT for smart home system, in: 2015 International Symposium on Consumer Electronics, ISCE, IEEE, 2015, pp. 1–2, http://dx.doi.org/10.1109/ISCE.2015.7177843.
[40] S. Hui, Z. Wang, X. Hou, X. Wang, H. Wang, Y. Li, D. Jin, Systematically quantifying IoT privacy leakage in mobile networks, IEEE Internet Things J. (2020) http://dx.doi.org/10.1109/JIOT.2020.3038639.
[41] S. Yu, G. Wang, X. Liu, J. Niu, Security and privacy in the age of the smart internet of things: An overview from a networking perspective, IEEE Commun. Mag. 56 (9) (2018) 14–18, http://dx.doi.org/10.1109/MCOM.2018.1701204.
[42] D. Geneiatakis, I. Kounelis, R. Neisse, I. Nai-Fovino, G. Steri, G. Baldini, Security and privacy issues for an IoT based smart home, in: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO, IEEE, 2017, pp. 1292–1297, http://dx.doi.org/10.23919/MIPRO.2017.7973622.
[43] P.C.M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, M. Atiquzzaman, A trustworthy privacy preserving framework for machine learning in industrial iot systems, IEEE Trans. Ind. Inf. 16 (9) (2020) 6092–6102, http://dx.doi.org/10.1109/TII.2020.2974555.
[44] F. Hussain, R. Hussain, S.A. Hassan, E. Hossain, Machine learning in IoT security: current solutions and future challenges, IEEE Commun. Surv. Tutor. (2020) http://dx.doi.org/10.1109/COMST.2020.2986444.
[45] M. Zheng, D. Xu, L. Jiang, C. Gu, R. Tan, P. Cheng, Challenges of privacy-preserving machine learning in iot, in: Proceedings of the First International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things, 2019, pp. 1–7, http://dx.doi.org/10.1145/3363347.3363357.
[46] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, A.V. Vasilakos, The quest for privacy in the internet of things, IEEE Cloud Comput. 3 (2) (2016) 36–45, http://dx.doi.org/10.1109/MCC.2016.28.
[47] J.C. Short, J.C. Broberg, C.C. Cogliser, K.H. Brigham, Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation, Organ. Res. Methods 13 (2) (2010) 320–347, http://dx.doi.org/10.1177/1094428109335949.
[48] K. Täuscher, R. Bouncken, R. Pesch, Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms, Acad. Manag. J. (2020) http://dx.doi.org/10.5465/amj.2018.0620, in press.
[49] D. Kim, K. Park, Y. Park, J.H. Ahn, Willingness to provide personal information: Perspective of privacy calculus in IoT services, Comput. Hum. Behav. 92 (2019) 273–281, http://dx.doi.org/10.1016/j.chb.2018.11.022.
[50] Purva Grover, Arpan Kumar Kar, Yogesh K. Dwivedi, Marijn Janssen, Polarization and acculturation in US election 2016 outcomes – Can Twitter analytics predict changes in voting preferences, Technol. Forecast. Soc. Change 145 (2019) 438–460, http://dx.doi.org/10.1016/j.techfore.2018.09.009.
[51] A. Alamsyah, W. Rizkika, D.D.A. Nugroho, F. Renaldi, S. Saadah, Dynamic large scale data on twitter using sentiment analysis and topic modeling, in: 2018 6th International Conference on Information and Communication Technology, ICoICT, 2018, pp. 254–258.
[52] A. Sarlan, C. Nadam, S. Basri, Twitter sentiment analysis, in: Proceedings of the 6th International Conference on Information Technology and Multimedia, 2014, pp. 212–216, http://dx.doi.org/10.1109/ICIMU.2014.7066632.
[53] J.R. Saura, D. Palacios-Marqués, A. Iturricha-Fernández, Ethical Design in Social Media: Assessing the main performance measurements of user online behavior modification, J. Bus. Res. 129 (2021) 271–281, http://dx.doi.org/10.1016/j.jbusres.2021.03.001.
[54] Muhammad Alam, Fazeel Abid, Cong Guangpei, L.V. Yunrong, Social media sentiment analysis through parallel dilated convolutional neural network for smart city applications, Comput. Commun. 154 (2020) 129–137, http://dx.doi.org/10.1016/j.comcom.2020.02.044.
[55] William M. Webberley, Stuart M. Allen, Roger M. Whitaker, Retweeting beyond expectation: Inferring interestingness in Twitter, Comput. Commun. 73 (2016) 229–235, http://dx.doi.org/10.1016/j.comcom.2015.07.016.
[56] Pardis Emami-Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, Norman Sadeh, Privacy expectations and preferences in an IoT world, 15, 2020.
[57] K. Yamazaki, R. Ueda, S. Nozawa, M. Kojima, K. Okada, K. Matsumoto, M. Ishikawa, I. Shimoyama, M. Inaba, Home-assistant robot for an aging society, Proc. IEEE 100 (8) (2012) 2429–2441, http://dx.doi.org/10.1109/JPROC.2012.2200563.
[58] A. Das, M. Degeling, N. Sadeh, Personalized privacy assistants for the internet of things: Providing users with notice and choice, IEEE Pervasive Comput. 17 (3) (2018) 35–46, http://dx.doi.org/10.1109/MPRV.2018.03367733.
[59] R. Roman, P. Najera, J. Lopez, Securing the internet of things, Computer 44 (9) (2011) 51–58, http://dx.doi.org/10.1109/MC.2011.291.
[60] Omar Alrawi, Chaz Lever, Manos Antonakakis, Fabian Monrose, SoK: Security evaluation of home-based IoT deployments, in: 2019 IEEE Symposium on Security and Privacy, SP, IEEE, San Francisco, CA, USA, 2019, pp. 1362–1380.
[61] Hang Hu, Limin Yang, Shihan Lin, Gang Wang, A case study of the security vetting process of smart-home assistant applications, in: IEEE Security and Privacy Workshops, SPW, 2020, p. 6, http://dx.doi.org/10.1109/SPW50608.2020.00029.
[62] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, X.S. Shen, Security and privacy in smart city applications: Challenges and solutions, IEEE Commun. Mag. 55 (1) (2017) 122–129, http://dx.doi.org/10.1109/MCOM.2017.1600267CM.
[63] Z. Zhang, R. Sun, C. Zhao, J. Wang, C.K. Chang, B.B. Gupta, Cyvod: a novel trinity multimedia social network scheme, Multimedia Tools Appl. 76 (18) (2017) 18513–18529, http://dx.doi.org/10.1007/s11042-016-4162-z.
[64] J.R. Saura, D. Ribeiro-Soriano, D. Palacios-Marqués, Setting B2B Digital Marketing in Artificial Intelligence-based CRMs: A review and directions for future research, Industrial Marketing Management 98 (2021) 161–178, http://dx.doi.org/10.1016/j.indmarman.2021.08.006.
[65] Denis Parra, Christoph Trattner, Diego Gómez, Matías Hurtado, Xidao Wen, Yu-Ru Lin, Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences, Comput. Commun. 73 (2016) 301–314, http://dx.doi.org/10.1016/j.comcom.2015.07.001.
[66] Prerna Mishra, Ranjana Rajnish, Pankaj Kumar, Sentiment analysis of Twitter data: Case study on digital India, in: 2016 International Conference on Information Technology (InCITe) - the Next Generation IT Summit on the Theme - Internet of Things: Connect Your Worlds, IEEE, Noida, 2016, pp. 148–153.
[67] K.L. Liu, W.J. Li, M. Guo, Emoticon smoothed language models for twitter sentiment analysis, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 26, No. 1, 2012, July.
[68] Marijn Janssen, Sunil Luthra, Sachin Mangla, Nripendra P. Rana, Yogesh K. Dwivedi, Challenges for adopting and implementing IoT in smart cities: An integrated MICMAC-ISM approach, Internet Res. 29 (6) (2019) 1589–1616, http://dx.doi.org/10.1108/INTR-06-2018-0252.
[69] N. Komninos, E. Philippou, A. Pitsillides, Survey in smart grid and smart home security: Issues, challenges and countermeasures, IEEE Commun. Surv. Tutor. 16 (4) (2014) 1933–1954, http://dx.doi.org/10.1109/COMST.2014.2320093.
[70] Tommaso Pecorella, Laura Pierucci, Francesca Nizzi, Network sentiment framework to improve security and privacy for smart home, Future Internet 10 (12) (2018) 125, http://dx.doi.org/10.3390/fi10120125.
[71] I. Pollach, Taming textual data: The contribution of corpus linguistics to computer-aided text analysis, Organ. Res. Methods 15 (2) (2012) 263–287, http://dx.doi.org/10.1177/1094428111417451.
[72] A.F. McKenny, H. Aguinis, J.C. Short, A.H. Anglin, What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis, J. Manag. 44 (7) (2018) 2909–2933, http://dx.doi.org/10.1177/0149206316657594.
[73] X. Xuanyang, G. Yuchang, W. Shouhong, L. Xi, Computer aided detection of SARS based on radiographs data mining, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, 2006, pp. 7459–7462, http://dx.doi.org/10.1109/IEMBS.2005.1616237.
[74] N. Leon, The future of computer-aided innovation, Comput. Ind. 60 (8) (2009) 539–550, http://dx.doi.org/10.1016/j.compind.2009.05.010.
[75] K. Krippendorff, Reliability in content analysis: Some common misconceptions and recommendations, Hum. Commun. Res. 30 (3) (2004) 411–433, http://dx.doi.org/10.1007/s11846-019-00333-8.
[76] F.N. Ribeiro, M. Araújo, P. Gonçalves, M.A. Gonçalves, F. Benevenuto, Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods, EPJ Data Sci. 5 (1) (2016) 1–29, http://dx.doi.org/10.1140/epjds/s13688-016-0085-1.
[77] S. Vijayarani, R. Janani, Text mining: open-source tokenization tools-an analysis, Adv. Comput. Intell. Int. J. 3 (1) (2016) 37–47.
[78] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, L. Zhao, Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, Multimedia Tools Appl. 78 (11) (2019) 15169–15211, http://dx.doi.org/10.1007/s11704-009-0062-y.
[79] T.A. Rana, Y.N. Cheah, S. Letchmunan, Topic modeling in sentiment analysis: A systematic review, J. ICT Res. Appl. 10 (1) (2016) http://dx.doi.org/10.5614/itbj.ict.res.appl.2016.10.1.6.
[80] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Sage Publications, 2018, http://dx.doi.org/10.1111/j.1468-4446.2007.00153_10.x.
[81] M. De Choudhury, Y.R. Lin, H. Sundaram, K.S. Candan, L. Xie, A. Kelliher, How does the data sampling strategy impact the discovery of information diffusion in social media? in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 4, No. 1, 2010.
[82] J. Hirschberg, C.D. Manning, Advances in natural language processing, Science 349 (6245) (2015) 261–266, http://dx.doi.org/10.1126/science.aaa8685.
[83] H. Bhavsar, R. Manglani, Sentiment analysis of Twitter data using Python, Int. Res. J. Eng. Technol. 6 (3) (2019) 510–527.

[84] Saura, Bennett, A three-stage method for data text mining: Using UGC in business intelligence analysis, Symmetry 11 (4) (2019) 519, http://dx.doi.org/10.3390/sym11040519.
[85] B.N. Hiremath, M.M. Patil, Enhancing optimized personalized therapy in clinical decision support system using natural language processing, J. King Saud Univ.-Comput. Inf. Sci. (2020) http://dx.doi.org/10.1016/j.jksuci.2020.03.006.
[86] N. Hardeniya, J. Perkins, D. Chopra, N. Joshi, I. Mathur, Natural Language Processing: Python and NLTK, Packt Publishing Ltd, 2016.
[87] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis, Comput. Linguist. 35 (3) (2009) 399–433, http://dx.doi.org/10.1162/coli.08-012-R1-06-90.
[88] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of human language technology conference and conference on empirical methods in natural language processing, 2005, pp. 347–354, http://dx.doi.org/10.3115/1220575.1220619.
[89] D. Sarkar, S. Markovski, M. Gusev, D. Tomp, S. Muravyov, A. Filchenkov, A. Parundekar, S. Elias, A. Ashok, S. Sujitparapitaya, A. Shirani, M. Roldan, V. Bonta, N. Kumaresh, N. Janardhan, S. Mishra, A. Pappu, N. Bhamidipati, L. He, K. Zheng, A. Bandi, A. Fellah, Inferring advertiser sentiment in online articles using Wikipedia footnotes, Adv. Intell. Syst. Comput. 2 (2019) 1224–1231, http://dx.doi.org/10.29007/kzk1.
[90] D. Griol, J.M. Molina, Z. Callejas, Combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances, Neurocomputing (2017) 1–9, http://dx.doi.org/10.1016/j.neucom.2017.01.120.
[91] J. Karoui, F.B. Zitoune, V. Moriceau, SOUKHRIA: towards an irony detection system for arabic in social media, Procedia Comput. Sci. 117 (2017) 161–168, http://dx.doi.org/10.1016/j.procs.2017.10.105.
[92] L. AlSumait, D. Barbará, J. Gentle, C. Domeniconi, Topic significance ranking of LDA generative models, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, Berlin, Heidelberg, 2009, pp. 67–82.
[93] A. Agrawal, W. Fu, T. Menzies, What is wrong with topic modeling? And how to fix it using search-based software engineering, Inf. Softw. Technol. 98 (2018) 74–88, http://dx.doi.org/10.1016/j.infsof.2018.02.005.
[94] Y. Chen, H. Zhang, R. Liu, Z. Ye, J. Lin, Experimental explorations on short text topic mining between LDA and NMF based schemes, Knowl.-Based Syst. 163 (2019) 1–13, http://dx.doi.org/10.1016/j.knosys.2018.08.011.
[95] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (Jan) (2003) 993–1022.
[96] Z. Khan, N. Iltaf, H. Afzal, H. Abbas, DST-HRS: A topic driven hybrid recommender system based on deep semantics, Comput. Commun. (2020) http://dx.doi.org/10.1016/j.comcom.2020.02.068.
[97] T. Loughran, B. McDonald, Textual analysis in accounting and finance: A survey, J. Account. Res. 54 (4) (2016) 1187–1230, http://dx.doi.org/10.1111/1475-679X.12123.
[98] S. Pandey, S.K. Pandey, Applying natural language processing capabilities in computerized textual analysis to measure organizational culture, Organ. Res. Methods 22 (3) (2019) 765–797, http://dx.doi.org/10.1177/1094428117745648.
[99] W. Qi, R. Procter, J. Zhang, W. Guo, Mapping consumer sentiment toward wireless services using geospatial twitter data, IEEE Access 7 (2019) 113726-113739, http://dx.doi.org/10.1109/ACCESS.2019.2935200.
[100] A. Reyes-Menendez, J.R. Saura, B.T. Stephen, Exploring key indicators of social identity in the #metoo era: Using discourse analysis in UGC, Int. J. Inf. Manage. 54 (2020) 102129, http://dx.doi.org/10.1016/j.ijinfomgt.2020.102129.
[101] M.W. Wu, K.Y. Su, Corpus-based automatic compound extraction with mutual information and relative frequency count, in: Proceedings of Rocling VI Computational Linguistics Conference VI, 1993, pp. 207–216.
[102] G. Bouma, Normalized (pointwise) mutual information in collocation extraction, in: Proceedings of GSCL, 2009, pp. 31–40.
[103] S. Iyengar, G. Sood, Y. Lelkes, Affect, not ideologya social identity perspective on polarization, Public Opin. Q. 76 (3) (2012) 405–431, http://dx.doi.org/10.1093/poq/nfs038.
[104] M. Hoffman, F. Bach, D. Blei, Online learning for latent dirichlet allocation, Adv. Neural Inf. Process. Syst. 23 (2010) 856–864.
[105] P. Rayson, R. Garside, Comparing corpora using frequency profiling, in: The Workshop on Comparing Corpora, 2000, pp. 1–6, http://dx.doi.org/10.3115/1117729.1117730.
[106] D. Biber, If you look at . . . : Lexical bundles in university teaching and textbooks, Appl. Linguist. 25 (3) (2004) 371–405, http://dx.doi.org/10.1093/applin/25.3.371.
[107] T. McEnery, A. Hardie, The history of corpus linguistics, in: The Oxford Handbook of the History of Linguistics, Vol. 727, 2013, p. 745, http://dx.doi.org/10.1093/oxfordhb/9780199585847.013.0034.
[108] W. Zhou, Y. Jia, A. Peng, Y. Zhang, P. Liu, The effect of iot new features on security and privacy: New threats, existing solutions, and challenges yet to be solved, IEEE Internet Things J. 6 (2) (2018) 1606–1616, http://dx.doi.org/10.1109/JIOT.2018.284773.
[109] W. Kong, F. Qiao, Q. Wu, Real-manufacturing-oriented big data analysis and data value evaluation with domain knowledge, Comput. Statist. 35 (2) (2020) 515–538, http://dx.doi.org/10.1007/s00180-019-00919-6.
[110] J.S. Kumar, D.R. Patel, A survey on internet of things: Security and privacy issues, Int. J. Comput. Appl. 90 (11) (2014) http://dx.doi.org/10.5120/15764-4454.
[111] S.M. Mohammad, Security and privacy concerns of the 'Internet of Things'(IoT) in IT and its help in the various sectors across the world, Int. J. Comput. Trends Technol. 68 (4) (2020) Available at SSRN: https://ssrn.com/abstract=3630513.
[112] A. Arabo, I. Brown, F. El-Moussa, Privacy in the age of mobility and smart devices in smart homes, in: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, IEEE, 2012, pp. 819–826.
[113] J. Bugeja, A. Jacobsson, P. Davidsson, On privacy and security challenges in smart connected homes, in: 2016 European Intelligence and Security Informatics Conference, EISIC, IEEE, 2016, pp. 172–175, http://dx.doi.org/10.1109/EISIC.2016.044.
[114] H. Cai, B. Xu, L. Jiang, A.V. Vasilakos, IoT-based big data storage systems in cloud computing: perspectives and challenges, IEEE Internet Things J. 4 (1) (2016) 75–87, http://dx.doi.org/10.1109/JIOT.2016.2619369.
[115] P. Radanliev, D.C. De Roure, R. Nicolescu, M. Huth, R.M. Montalvo, S. Cannady, P. Burnap, Future developments in cyber risk assessment for the internet of things, Comput. Ind. 102 (2018) 14–22, http://dx.doi.org/10.1016/j.compind.2018.08.002.
[116] C. Patel, N. Doshi, Security challenges in IoT cyber world, in: Security in Smart Cities: Models, Applications, and Challenges, Springer, Cham, 2019, pp. 171–191, http://dx.doi.org/10.1007/978-3-030-01560-2.
[117] P.N. Mahalle, B. Anggorojati, N.R. Prasad, R. Prasad, Identity authentication and capability-based access control (iacac) for the internet of things, J. Cyber Secur. Mobil. (2012) 309–348.
[118] H. Sedjelmaci, S.M. Senouci, T. Taleb, An accurate security game for low-resource IoT devices, IEEE Trans. Veh. Technol. 66 (10) (2017) 9381–9393, http://dx.doi.org/10.1109/TVT.2017.2701551.
[119] S. Popereshnyak, O. Suprun, O. Suprun, T. Wieckowski, Iot application testing features based on the modelling network, in: IoT Application Testing Features Based on the Modelling Network. in 2018 XIV-Th International Conference on Perspective Technologies and Methods in MEMS Design, MEMSTECH, IEEE, 2018, pp. 127–131, http://dx.doi.org/10.1109/MEMSTECH.2018.8365717.
[120] F. Salamone, L. Belussi, C. Currò, L. Danza, M. Ghellere, G. Guazzi, et al., Integrated method for personal thermal comfort assessment and optimization through users' feedback, IoT and machine learning: a case study, Sensors 18 (5) (2018) 1602, http://dx.doi.org/10.3390/s18051602.
[121] M.A. Amanullah, R.A.A. Habeeb, F.H. Nasaruddin, A. Gani, E. Ahmed, A.S.M. Nainar, et al., Deep learning and big data technologies for IoT security, Comput. Commun. 151 (2020) 495–517, http://dx.doi.org/10.1016/j.comcom.2020.01.016.
[122] H. Lee, A. Kobsa, Confident privacy decision-making in IoT environments, ACM Trans. Comput.-Hum. Interact. 27 (1) (2019) 1–39, http://dx.doi.org/10.1145/3364223.
[123] N. Madaan, M.A. Ahad, S.M. Sastry, Data integration in IoT ecosystem: Information linkage as a privacy threat, Comput. Law Secur. Rev. 34 (1) (2018) 125–133, http://dx.doi.org/10.1016/j.clsr.2017.06.007.
[124] G. Ishmaev, The ethical limits of blockchain-enabled markets for private IoT data, Philos. Technol. 33 (3) (2020) 411–432, http://dx.doi.org/10.1007/s13347-019-00361-y.
[125] B. Farahani, F. Firouzi, V. Chang, M. Badaroglu, N. Constant, K. Mankodiya, Towards fog-driven IoT ehealth: Promises and challenges of IoT in medicine and healthcare, Future Gener. Comput. Syst. 78 (2018) 659–676, http://dx.doi.org/10.1016/j.future.2017.04.036.
[126] H. Khemissa, D. Tandjaoui, A lightweight authentication scheme for E-health applications in the context of internet of things, in: 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies, IEEE, 2015, pp. 90–95.
[127] P.H. Vilela, J.J. Rodrigues, P. Solic, K. Saleem, V. Furtado, Performance evaluation of a Fog-assisted IoT solution for e-Health applications, Future Gener. Comput. Syst. 97 (2019) 379–386, http://dx.doi.org/10.1016/j.future.2019.02.055.