International Journal of Information Management Data Insights 1 (2021) 100032

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

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

# Application of text mining in smart lighting literature - an analysis of existing literature and a research agenda

Atousa Zarindast*, Anuj Sharma, Jonathan Wood

*Department of Civil and Environment Engineering, Iowa State University, 2711 S Loop Dr, Ames, IA 50010, Ames, Iowa, USA*

ARTICLE INFO

*Keywords:*
Text mining
Research and development gap
Topic modeling
Smart housing
Aspect based sentiment analysis

ABSTRACT

There has always been a gap between research and engineering practices across almost all disciplines. The nature of these gaps are not identified; However, in order to be able to bridge these gaps a comprehensive understanding is required. In this study, in order to provide research agenda, we examine existing literature and users perception in smart lighting business. We consider two sources of data for this analysis 1-lighting research literature that reflects the research phase and 2-Amazon reviews about two smart lighting products that represent customers' perceptions reflecting the development phase. We use text mining and analyze both data sources by topic modeling and sentiment analysis. We conducted a probabilistic aspect based sentiment analysis. One of the results of this analysis showed that while the research community is drastically concerned about energy consumption, the end users were excited about color changing ability in smart lighting products.

## 1. Introduction

Smart cities and smart buildings are changing and shaping our civilization. Population growth pressures our needs to develop smart cities to bridge the economy and social life with focus on sustainability. Emerging technologies such as communication technologies (Internet of Things (IoT)) along with Artificial Intelligent (AI) enables such purposes. Combining AI with IoT has shown promising outcomes in the past (Chatterjee, Kar, & Gupta, 2018).

Research on developing smart buildings, making our everyday life easier and more efficient has become an interest recently (Al Dakheel, Del Pero, Aste, & Leonforte, 2020). Buildings are among the highest energy consumer compared to industry and transportation (Building Energy Data Book, 2011; Mishra, Irwin, Shenoy, Kurose, & Zhu, 2013). According to the US Department of Energy (DOE), 70% of overall electricity usage in the US comes from buildings (DOE Buildings Energy Data Book, 2009). A small improvement in efficiency of energy consumption, if adopted widely, has a high potential to impact the overall energy consumption. Moreover, 70% of buildings' energy usage is in form of electricity (Mishra et al., 2013). Therefore, energy consumption of buildings is a crucial issue that has attracted significant attention (Moadab, Olsson, Fischl, & Aries, 2021; Wagiman, Abdullah, Hassan, & Radzi, 2021). Smart cities and smart buildings are path toward a more sustainable life style (Mishra et al., 2013; Shamsuzzoha, Niemi, Piya, & Rutledge, 2021).

Smart lighting can potentially advance building management and energy efficiency (Chew, Karunatilaka, Tan, & Kalavally, 2017), and suitable lighting can also help with human psychology (Knez & Kers, 2000). In recent years, connectivity of devices and buildings called Internet of Things (IoT) has increased drastically (Harkare, Potdar, Mishra, Kekre, & Harkare, 2021; Le, Le Tuan, & Tuan, 2019; Minoli, Sohraby, & Occhiogrosso, 2017). Building management systems have become connected and intelligent. However, the adoption of smart home technologies is limited (Shin, Park, & Lee, 2018). Lighting is a one of the major energy consumer in buildings (Nicol, Wilson, & Chiancarella, 2006) and accounts for 5–15% of total electric energy consumption (Ryckaert, Lootens, Geldof, & Hanselaer, 2010). Increasing the adoption rate of smart technologies would result in energy consumption reduction. By understanding customers perception, and aligning those concerns with research communities and industry practitioners, we can potentially increase the adoption rate of smart lighting products (Chew et al., 2017; Zarindast, Wood, & Sharma, 2021). As a result, reduce the energy consumption and progress toward a more sustainable and efficient smart building management system within the smart city context.

One of the potentials of smart lighting business is the ability to save energy (Chew et al., 2017). Hence, considerable research have focus on energy saving through smart lighting systems design, development, and planning (Moadab et al., 2021; Wagiman et al., 2021; Xu et al., 2017). Energy is an important factor in developing lighting systems, but to make this newly developed system more attractive for average user, several other features could be embedded in the system. Results of

* Corresponding author.
*E-mail address:* Atousaz@iastate.edu (A. Zarindast).

https://doi.org/10.1016/j.jjimei.2021.100032
Received 9 June 2021; Received in revised form 11 August 2021; Accepted 11 August 2021

this analysis can further help to understand users' perceptions, needs, and concerns which can be useful in adoption of these product.

Lack of prior studies exploring the research and development gap might be due to the unavailability of data or tools required to do such analyses. Currently, online reviews are available and have been analyzed for numerous purposes including sentiment analysis (Jagdale, Shirsat, & Deshmukh, 2019), product recommendation (Yun, Hooshyar, Jo, & Lim, 2018), and service quality examination (Lucini, Tonetto, Fogliatto, & Anzanello, 2020). In addition, advanced text mining tools allow researchers to explore unstructured data in an efficient manner. Recently, text mining has been explored for different applications in innovative manners. For instance Jung & Lee (2020) utilized text mining to find similar experiments derived from past experiments in green building designs in order to reuse suggested solutions for new similar cases. Krallinger & Valencia (2005) utilized text mining to retrieve information about genes, proteins, and their relationship.

Topic modeling is one the power full techniques in text mining (Jelodar et al., 2019) with many application in various fields such as marketing (Amado, Cortez, Rita, & Moro, 2018), software engineering (Gethers & Poshyvanyk, 2010), political science (Greene & Cross, 2015), and medical science (Porturas & Taylor, 2020). There are various topic models such as Latent semantic analysis (LSA),Probabilistic latent semantic analysis (PLSA), Latent Dirichlet allocation (LDA), correlated topic model (CTM), etc. LDA is a model that improves exchangeability of words and documents compared to LSA and PLSA (Alghamdi & Alfalqi, 2015). Moreover, it is flexible enough to undertake rich analysis (Blei, 2012). LDA is one of well known and highly used algorithms in topic modeling. Recently, Yun (2020) reviewed trends of research in the field of physics education to identify the status of research and help researcher in this area using LDA and topic modeling. Choi, Lee, & Sohn (2017) used LDA to discover topics discussed in several different documents (e.g. Journal articles, books) to examine academic research and offer future direction. Yun & Geum (2020) used LDA for automated patent classification.

In this study, we analyzed possible gaps between research and development and provide research agenda, by utilizing the online reviews and peer reviewed journal articles. We used smart lighting as an application of the proposed framework and analyzed the users perception and compared it with the state of the art literature. We used advanced text mining tools such as topic modeling and LDA to perform this analysis. Through this study, we can contribute to our understanding of possible research and development gap, particularly in smart lighting business. We can create a path toward further analysis in this area and as a result help in bridging any possible gap. It will also provide insights that help in analyzing users' perceptions and needs, and by aligning research and development with those concerns, it can help increase the adoption rates of smart lighting products.

We analyzed the household users' perceptions, particularly considering two main smart lighting products from two major companies in the smart lighting business.This study has three major impacts: 1- it contributes to our understanding of possible research and development gap, particularly in smart lighting business; 2- it creates a path toward further analysis in this area and provide research agenda; 3 - it provides insights that help in analyzing users' perceptions and needs, and by aligning research and development with those concerns, it can help increase the adoption rates of smart lighting products. The other section of the article are structured as follow: we present methodology in Section 2. Results are analyzed in Section 3 and discussion is in Section 4. Conclusion and future work are presented in Section 5.

## 2. Methodology

The first step in reducing the gap between research and development is identifying and understanding them. We illustrated possible gaps in research and development in production platform of smart lighting business. We used two sources of data for this analysis; namely peer reviewed articles in lighting research for the research phase and customer reviews of two major smart lighting products for the development phase. In order to analyze these two data corpus, we utilized text mining and topic modeling. Fig. 1 shows proposed framework for this analysis. First, we start by Latent Dirichlet allocation (LDA) topic modeling and find the per-word-per-topic and per-document-per-topic probabilities. Manually reviewing the top words and documents in each topic would help in labeling the extracted topics and their classes. Second, these probabilities were used to classify documents and words to the most probable topic. Finally, we examined the online review sentiments toward each specific topic. Fig. 2 shows the process framework used in this analysis. The process is explained with more detail in following sections:

### 2.1. Data gathering

#### 2.1.1. Peer reviewed journal articles data retrieval

The first step was data gathering and prepossessing. Different combinations of following terms: 'lighting', 'light, 'systems', 'control', 'smart' was used in a query search through "https://www.scopus.com/ " search engine, with following subject areas 'Engineering', 'Computer science', 'Energy',Environment', 'Social science' that resulted in 300 peer reviewed journal articles' abstract. We limited our search to English language journal articles published during years 2010 and 2021.

#### 2.1.2. Smart lighting products descriptions and data retrieval

On the other hand, we scrapped reviews published in www.amazon.com related to two major smart lighting products 1- Philips Hue White and Color Ambiance A19 60W Equivalent LED Smart Bulb Starter Kit product, and 2-Sengled Smart Light Bulbs product. We used R programming language and "rvest" package to scrape the data from amazon review pages. In this scrapping process along with the review text, title, date, author, location, stars, and review format were also collected. In scrapping process total of 1031 global reviews were retrieved on 11/01/2020 for Philips Hue and 52 global reviews were retrieved on 11/01/2020 for Sengled smart bulbs. We limited our analysis to reviews located in United states and this resulted in 750 reviews for Philips and 50 reviews for Sengled. After collecting the data, all letters were converted to lower case. Moreover, stop words, punctuation, and numbers were removed. We performed selective manual stemming for topic modeling and frequency analysis, for instance plural words were converted to singular and unnecessary words such as "get", "can", "it.", "one", "just", "set", "will", 'also', 'even', 'really', 'hub', 'bulb', 'bulbs', 'it', 'got', 'two', 'much', 'find', 'them.', 'back', 'recommend', 'now' were manually selected and removed from the corpus.

Figure 3 shows the underlying smart lighting products in this study. These two products were selected as they were the major known smart lighting products available. Both products have a hub (bridge) that communicates with the lights. They work with Amazon Alexa and Google Home voice commands to turn on/off the lights and adjust the color. These products provide an app to control the lights and colors.

### 2.2. Gap analysis and research agenda

In this study, we used two publicly available data sources. The aim of this study was to examine any possible gap between research and product development along with customers' perception and provide research agenda. We aim to show the differences and similarities between the main concerns of research society compared to concerns and perceptions of actual users. We illustrated this comparison between the research community and product development through examining peer reviewed journal articles and smart lighting products online reviews. To do so, we utilized text mining tools and implemented topic modeling, bigram/document classification, and aspect based sentiment analysis.

#### 2.2.1. Topic modeling

Enabling computers to obtain meaning from unstructured text documents written by humans is a very challenging research in computer sci-

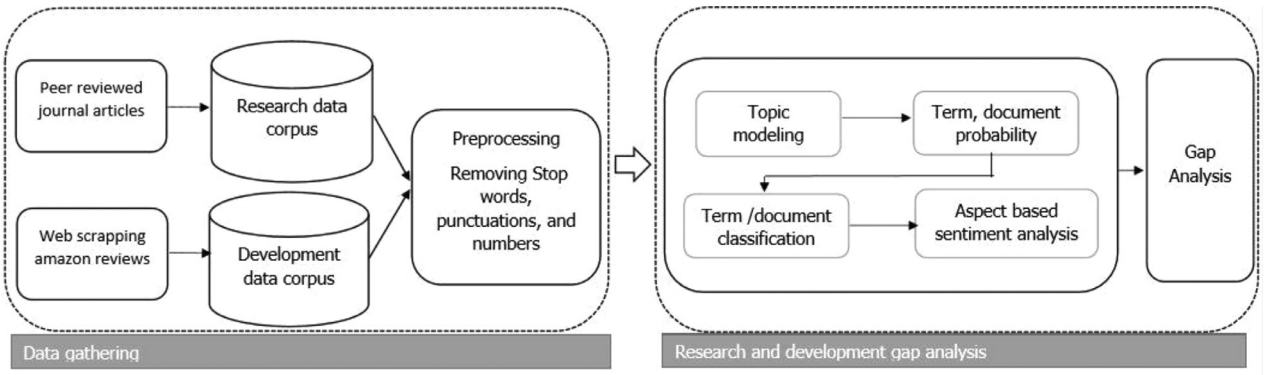



**Fig. 1.** Methodology.

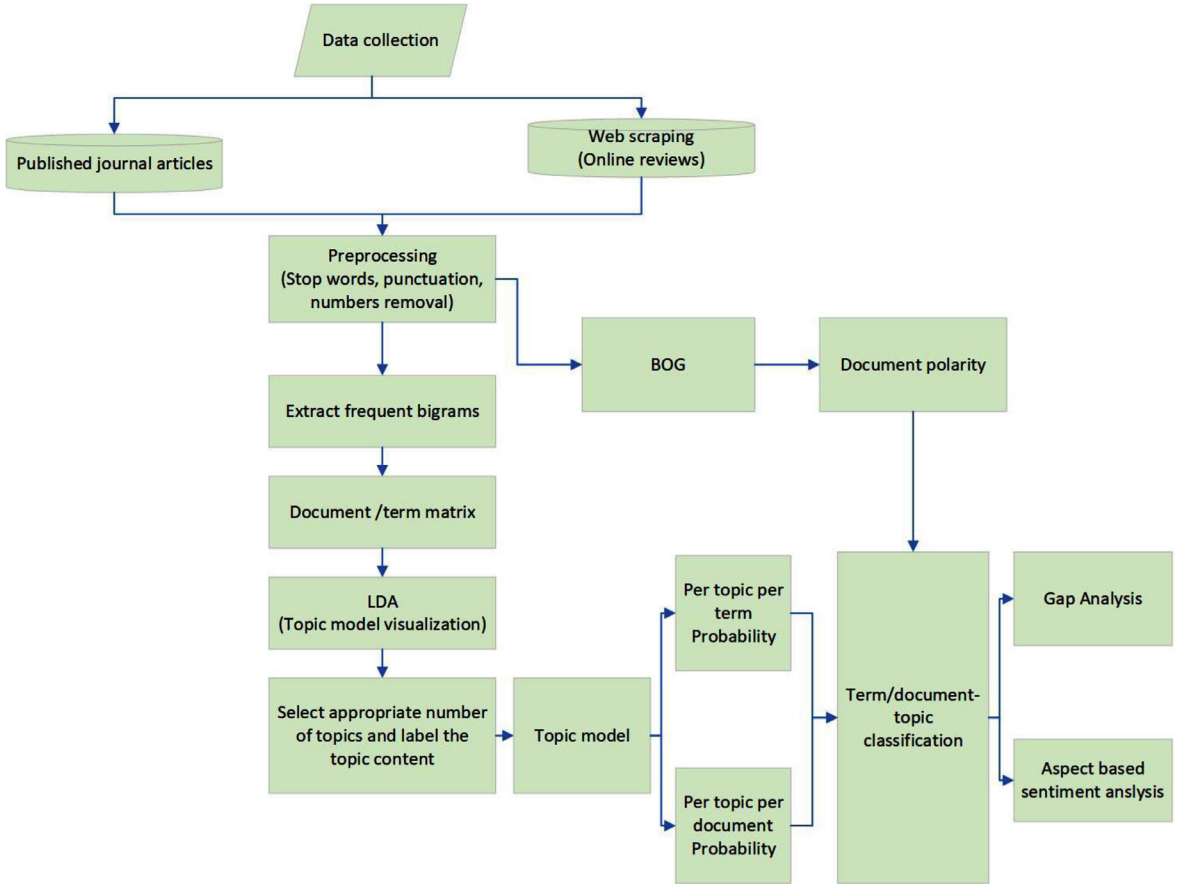



**Fig. 2.** Flow chart framework of the analysis.

ence. Hierarchical probabilistic models are used to find pattern of words in document collections and are called "topic models" (Alghamdi & Alfalqi, 2015). Topic modeling identifies a probabilistic procedure through which the document can be generated from. The main idea is to discover word-use patterns and connect documents sharing similar patterns. The generated topic is a probability distribution over words. There are various topic models such as Latent semantic analysis (LSA),Probabilistic latent semantic analysis (PLSA), Latent Dirichlet allocation (LDA), correlated topic model (CTM). LDA is a model that improves exchange ability of words and documents compared to LSA and PLSA (Alghamdi & Alfalqi, 2015) and it is flexible enough to undertake rich analysis (Blei, 2012). LDA is widely used for different applications. Recently Yun (2020) reviewed trends of research in the field of physics education to identify the status of research and help researcher in this area using LDA and topic modeling.

Lets define notations and definitions used in topic modeling in this study: A term is a unit of discrete data from our corpus. A document is a vector sequence of N terms denoted by $T = (t_1, t_2, \ldots t_N)$ and $t_n$ is the nth term in the vector sequence. Corpus is collection of M documents defined by $C = (T_1, T_2, \ldots T_M)$. Latent Dirichlet allocation (LDA) is a probabilistic model in which each topic is modeled by a distribution over terms. $\theta$ has a Dirichlet distribution $Dir(\alpha)$ with following probability density:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \ldots . \theta_k^{\alpha_k - 1} \quad (1)$$

Where $\alpha$ is a vector with length k and $\alpha_i > 0$, and $\gamma$ is the Gamma function and Z $\sim$ Multinomial ($\theta$). Given parameters $\alpha$ and $\beta$ and The joint distribution of a topic with set of N topics $z$ and terms $T$ is defined


(a) Sengled Smart Light Bulbs - color (b) Philips Hue White and Color Ambiance changing light bulbs A19 60W Equivalent LED Smart Bulb Starter Kit

**Fig. 3.** Smart lighting products.

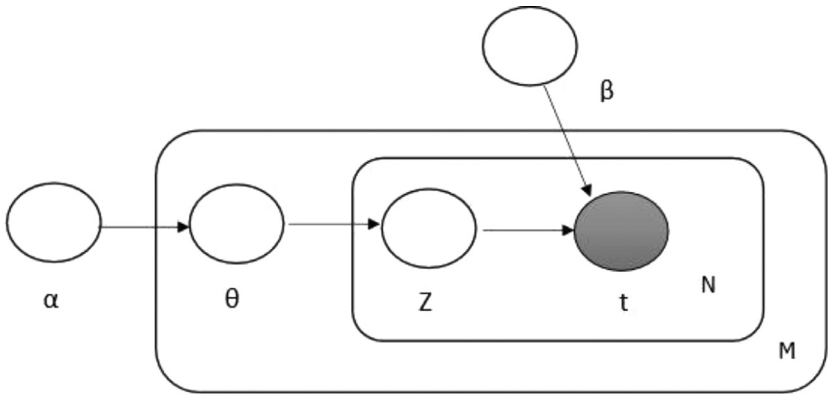



**Fig. 4.** Graphical modeling representation in LDA. Outer box represents document and inner box represents terms within a document (Blei et al., 2003).

bellow:

$$p(\theta, z, T|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(t_n|z_n, \beta)d\theta \quad (2)$$

Integrating over $\theta$ and summing over z, marginal distribution of a document is as follow:

$$p(T|\alpha, \beta) = \int p(\theta|\alpha)\left(\prod_{n=1}^{N} p(z_n|\theta)p(t_n|z_n, \beta)\right)d\theta \quad (3)$$

Taking product of marginal probabilities of each document in a corpus we have probability of a corpus as follow:

$$p(C|\alpha, \beta) = \prod_{d=1}^{M}\int p(\theta_d|\alpha)\left(\prod_{d=1}^{N_d}\sum_{z_{dn}} p(z_{dn}|\theta_d)p(t_{dn}|z_{dn}, \beta)\right)d\theta_d \quad (4)$$

A graphical representation of LDA model is presented in Fig. 4 and there are three levels on the LDA representation. The prarameters $\alpha$ and $\beta$ are corpus level parameters (C), $\theta_d$ is document level variable, $z_{dn}$ and $t_{dn}$ are term level variables. For detailed information on LDA methodology, we refer readers to Blei, Ng, & Jordan (2003).

To facilitate the E-step, posterior distribution $p(\theta, z|T, \alpha, \beta)$ was replaced by variational distribution $q(\theta, z|T, \gamma, \phi)$. For a given document T, the variational parameters $\gamma$ and $\phi$ are determined as follow and $D_{KL}$ denotes the Kullback-Leibler (KL) divergence (Blei et al., 2003) :

$$(\gamma^*, \phi^*) = argmin_{\gamma,\phi} D_{KL}(q(\theta, z|\gamma, \phi)||p(\theta, z|w, \alpha, \beta)) \quad (5)$$

### 2.2.2. Bigram, document classification

In order to build a topic model from our corpus after prepossessing steps we first identified frequent bigrams by limiting the frequency to three. Later we adopted LDA on resulted bigrams. Interactive visualization is a method for selecting the number of topics and interpreting the topic context which has been utilized by studies in topic modeling (Alam, Ofli, & Imran, 2020; Mortenson & Vidgen, 2016). Therefore, we used LDAVis package Sievert & Shirley (2014) that gives an overview of all the topics and terms as shown in Fig. 5. We experiment with different values of k and finalized the number of meaning full topics. We removed some of words in the data set that may cause discrimination, for instance hue and philips.

We used a semi supervised method to classify each document/bigram to one the identified topic in our topic modeling. For this purpose, we extracted per-topic-per-bigram probabilities from the LDA model.

Table 1 shows per topic per bigram probabilities for few terms. For instance there is a 0.00603 probability that term 'Electrical lighting' was generated from topic number 2 while there is 0.0257 probability that 'Energy demand' was generated from topic number 3.

Each bigram has been classified to topics based on highest probabilities. In order to understand the concept and context of each of these topics ([1–5]), we used interactive visualization and reviewed the relevant and high frequent bigrams in each topic. It is notable since the total number of reviews related one of the product (Sengled) was very low,

**Table 1**
Per-topic-per-bigram probabilities (partial).

| Topic | Bigram | beta |
|---|---|---|
| 2 | Electrical lighting | 0.00603 |
| 5 | Collected data | 0.00560 |
| 1 | Lighting conventional | 0.00525 |
| 2 | Control shading | 0.00521 |
| 2 | Saving energy | 0.00521 |
| 5 | Lighting natural | 0.00506 |
| 3 | Energy demand | 0.00489 |

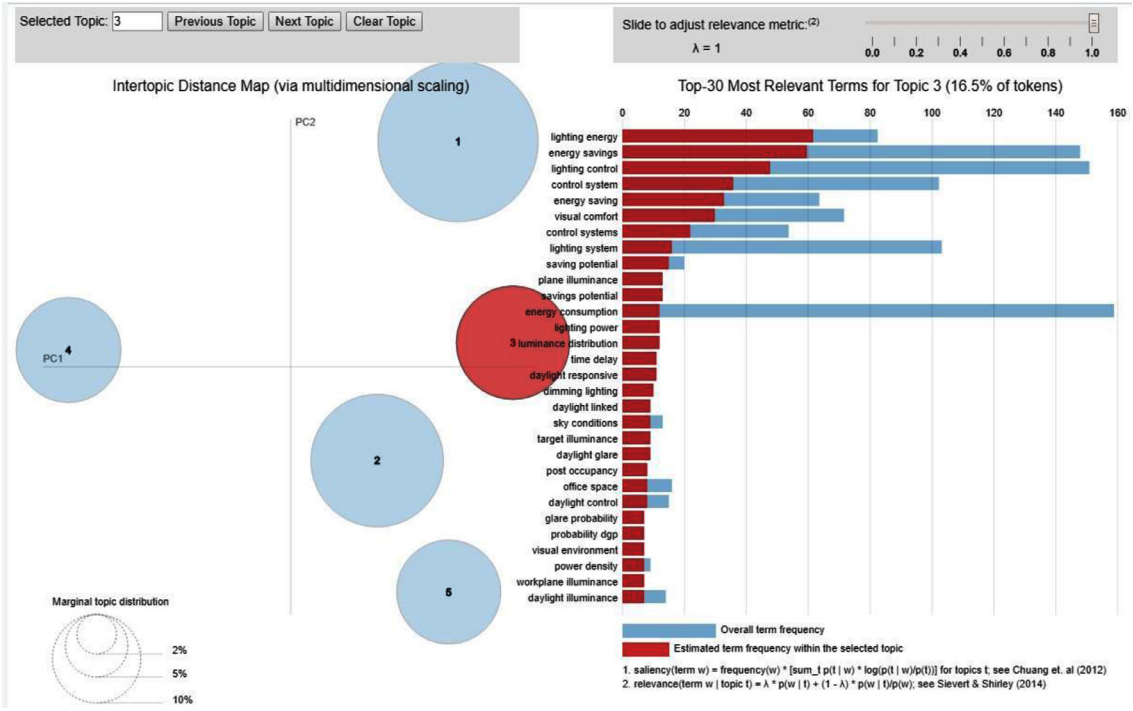



**Fig. 5.** Interactive visualization of topic models .

the frequency of bigrams was insufficient. As a result instead of frequent bigrams, terms were used as input of LDA.

This visualization helped us in understanding the four topics that were extracted from the Amazon reviews. Most common words in topic 1 included “google home”, “smart lights”, “white lights”, “soft white”, “bright white”, “color combination”, and “color changing which suggested that this topic represents ***color and changing colors***. Similarly, most common and relevant words in topic 2 included “starter kit”, “google assistant”, “turn” and “apple homekit”, “party app”, “alexa app”, “super easy”, “amazon alexa”, and “app makes” suggesting that, this topic was about ***ease of use***. LDA also produces per-document-per-topic probabilities, and each document was classified to each topic based on the highest probability.

### *2.2.3. Aspect based sentiment analysis*

Discovering ideas, emotions, and opinions for large amount of textual data is extremely difficult. Therefore, natural language processing provides a tool to extract customers opinion from large amount of unstructured data. Sentiment analysis refers to automating the text emotion detection process. We used “AFINN” dictionary for sentiment analysis; this dictionary includes a list of words with assigned sentiment values between (−5,5), representing negative and positive polarity. For this section using predefined dictionary, the overall sentiment value in each document is calculated and polarity is determined. Aspect based sentiment analysis is based on polarity of the document and the document-topic classification.

## 3. Results and discussion

In this study we aim to understand any possible gap between research and development phase and provide research agenda. To do so we utilized text mining on two publicly available data sources, 1- peer reviewed journal articles representing research phase and 2- online users reviews representing development phase. It is noteable that analyzing online reviews would illustrate users perceptions and concerns regarding what is being produced and used. This comparison is based on the topics that are discussed in these two data sources and shows the differences / similarities of concerns in these two communities, particularly in smart lighting business. We first started by extracting per-bigram-per-topic and per-document-per-topic probabilities. Moreover, reviewing the most frequently used bigram and sample documents in each topic resulted in topic labeling shown in Table 2.

After prepossessing and calculating document term matrix, term frequency was obtained and used for word cloud visualization presented in Fig. 6. This visualization shows what words and how frequently they have been mentioned in the corpus. As shown in Fig. 6, bigger terms represents more frequently used words.

### *3.1. Topic modeling*

Text mining can assist in automating the textual documents analysis. Topic modeling projects a document to multiple topics. We used a probabilistic method (LDA) to further classify terms and documents contributing to each topic. Reviewing top words (shown in Fig. 5) in each extracted class (topic) assisted in interpretation and labeling each class (topic). Figure 7 shows topics extracted from online reviews related to two major smart lighting products namely Philips Hue and Sengled smart light bulbs. Extracted topics for these two products were similar. Customers were mainly focused on ‘ease of use’, the color changing feature of bulbs, having control over the lights and system, with multiple command sources such as Google Home, Amazon Alexa, and by mobile application. Finally, they were focused on smartness of lighting systems which enables users with flexible advantages such as setting up rules and routines. As shown in Fig. 7, color, controllable system, and ease of use are among the most frequently referenced topics in both products.

***Color*** was the most common aspect in online reviews for the Philips Hue lighting product and it is not surprising as Philips Hue supports multiple predefined, as well as user defined, color ambiances called “scenes”. These scenes create a specific ambiance in the room that is adjustable by the Hue mobile app depending on occupants’ mood. Therefore, as this feature is relatively new it is understandable that users were excited to explore and comment about color changing feature. For instance, one user states that “*..Over all the lights are nice. The color is great, 3rd gen is the charm with every color emitted being very close to what I picked*

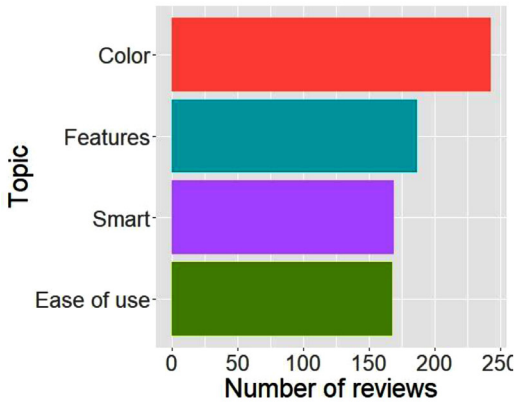
**Table 2**
Extracted topics.

| Source | Peer reviewed journal articles' abstract | Philips Hue online reviews | Sengled-color bulb online reviews |
|---|---|---|---|
| Topics | Energy | Colors | Control |
| | Comfort | Easy to use | Easy to use |
| | Smart | Features and control | color |
| | Building & infrastructure | Smart | - |
| | Occupancy based control systems | - | - |
| | | - | - |

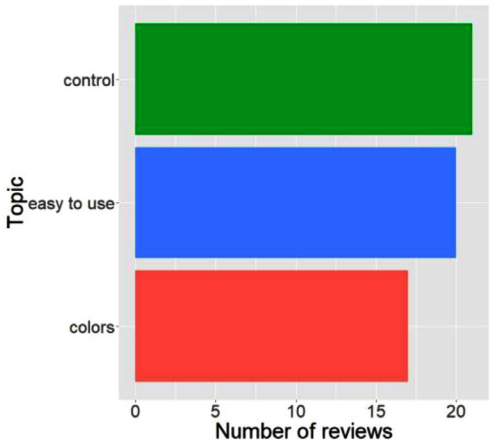

(a) Philips Hue starter kit

(b) Peer reviewed journal articles

**Fig. 6.** World cloud.




(a) Philips Hue White and Color Ambiance A19 60W Equivalent LED Smart Bulb Starter Kit




(b) Sengled Smart Light Bulbs - color changing light bulbs

**Fig. 7.** Number of reviews in each topic: Smart lighting products .

*in-app. I saw no flicker when dimming, or while dimmed, and the bulbs updated with the new settings within a second. They even fit in a lamp that has that wire loop that clips to bulb.*" Another user commented that "*.....The app has several color "scenes" to choose from... or simply use the color palette to create your own color combinations."*

***Control and features*** having control over the system was one the most important and frequently mentioned topics in online reviews. For instance in one review user commented that :"*..This is a safety practice and there's no way around it without customizing a timer or programmatic routine yourself. Additional accessories allow you to control the lights' settings without using your phone.*". This shows how different features such as buttons, switches, and sensors enables user to control the light and color in different ways.

Another aspect was ***ease of use***, as it is very important for users to be able to easily and quickly set up the system. For instance one user commented "*..The hub/lights/smartphone/Alexa setup sounds daunting but is actually pretty straight forward*" or another user commented "*..Setup was incredibly easy. Within minutes I had the 4 bulbs installed. Syncing with the app was a breeze.*", or in another review '*..Super easy to "install", easy to integrate to Home Kit and just overall easy to use."*

Finally, ***smartness*** was one of the categories that users were very enthusiastic about. For instance, "*..Google as been working without issues to turn lights on and off and the app works every time to make color or dimming adjustments. You can setup light schedules, etc as well*" or "*..Combining this with a tap switch, Alexa, and the Philips Hue app makes this the killer smart light setup.*". Overall, results are similar in these products and do

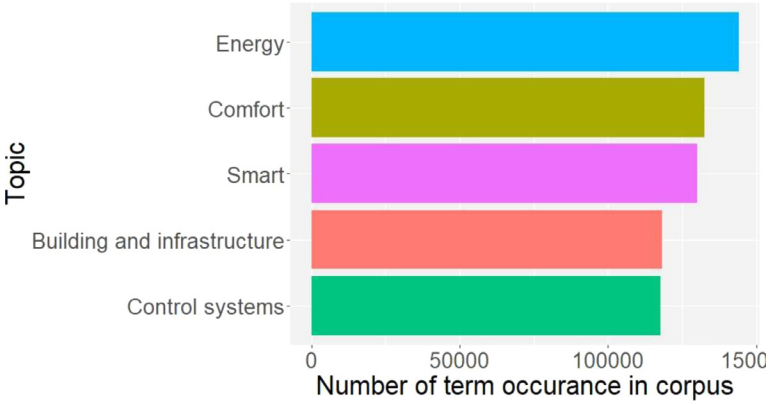
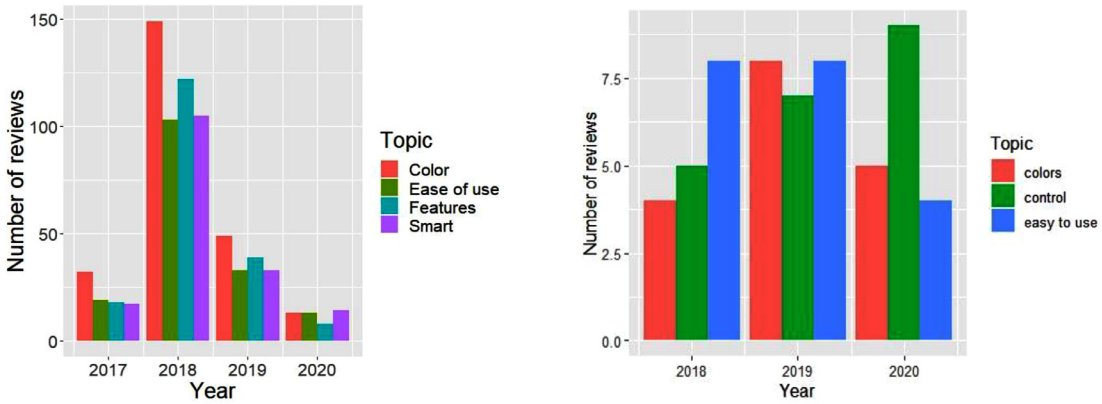



**Fig. 8.** Number of topic occurrence in peer reviewed articles abstracts.




(a) Philips Hue White and Color Ambiance A19 60W Equivalent LED Smart Bulb Starter Kit

(b) Sengled Smart Light Bulbs - color changing light bulbs

**Fig. 9.** Number of reviews yearly basis: Smart lighting products.

not appear to be very sensitive to particular product's review. Moreover, results are not impacted by differences in products.

On the development front shown in Fig. 8, consumers were concerned about ease of use, control over the system, smartness, and features. While on the research front, the research community mainly discussed topics such as occupancy based light control system development, the energy consumption, managing and energy reduction resulting from light control systems. The main root of all research studies was about energy consumption, meanwhile, the discussion around the system and/or infrastructure needed for light control system grew from the this root. After energy consumption/system control/infrastructure, user comfort was the most cited topic in research community. The obvious gap between research community and development and users perception is the energy consumption topic. While energy consumption and reduction is the major concern for research community, consumers do not seem to care about how the product consumes energy, rather they were more excited about color changing ability, ease of use and having control over the system.

Figure 9 shows number of reviews related to each aspect on a yearly basis. In Fig. 9.a, the jump in 2018 is not surprising as Philips Hue starter kit product, was introduced in August 2017 and it should be in maturity phase in year 2018. As a result, people are more engaged in interacting, exploring, and commenting about different aspects and product features. As shown, color is the most exciting topic for users in 2018 followed by the ease of use and control capability. After maturity phase, there would be a steady adaptation for users that has been reflected by reduction in number of reviews in 2019 and 2020. On the other hand, for Sengled product shown in Fig. 9.b, ease of use and control were among most cited topics throughout the analysis period.

While users are excited about exploring different features in smart lighting products, research community was paying attention to energy consumption, shown in Fig. 10. As energy consumption reduction was introduced with automated lighting control systems and technologies, the discussion around occupancy based control systems and required infrastructures was bold as well (Caicedo, Pandharipande, & Leus, 2011; Gunay, O'Brien, Beausoleil-Morrison, & Gilani, 2017; Peruffo, Pandharipande, Caicedo, & Schenato, 2015). While the major concern for research community was energy consumption, given the importance of users satisfaction and comfort, research topic considering both comfort and light control system in same system got bold starting 2012 (Heydarian, Carneiro, Gerber, & Becerik-Gerber, 2015; de Korte et al., 2015). The Intelligence of the system is still in infancy for research community and got a lot of attentions in recent years.

### 3.2. Aspect based sentiment analysis

Aspect based sentiment analysis is examination of emotions with regard to each topic. Figure 11 illustrates users perception toward topics that were discussed in reviews regarding smart lighting business. Overall users were excited about *color* aspect in positive attitude. This is understandable as changing color is one of the main features differentiating smart lighting products with other light products. The ability to change the color of light based on users' mood or theme of the environment is a key feature that created a lot of attention. For instance a user commented that "*..As far as the lights, they're phenomenal. Good brightness, an awesome range of whites, and the colors are very vibrant. Allows my girlfriend and I to find a balance of white that doesn't make me feel like I live in a Walmart (cold white), and doesn't make her feel like she lives in a bowl*

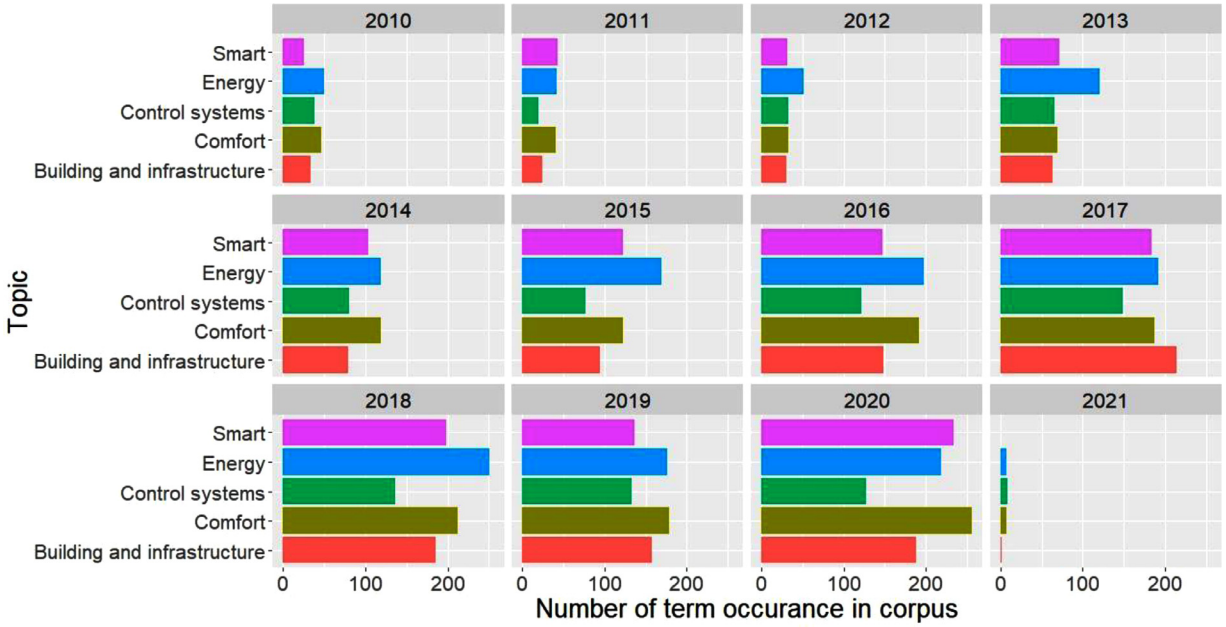
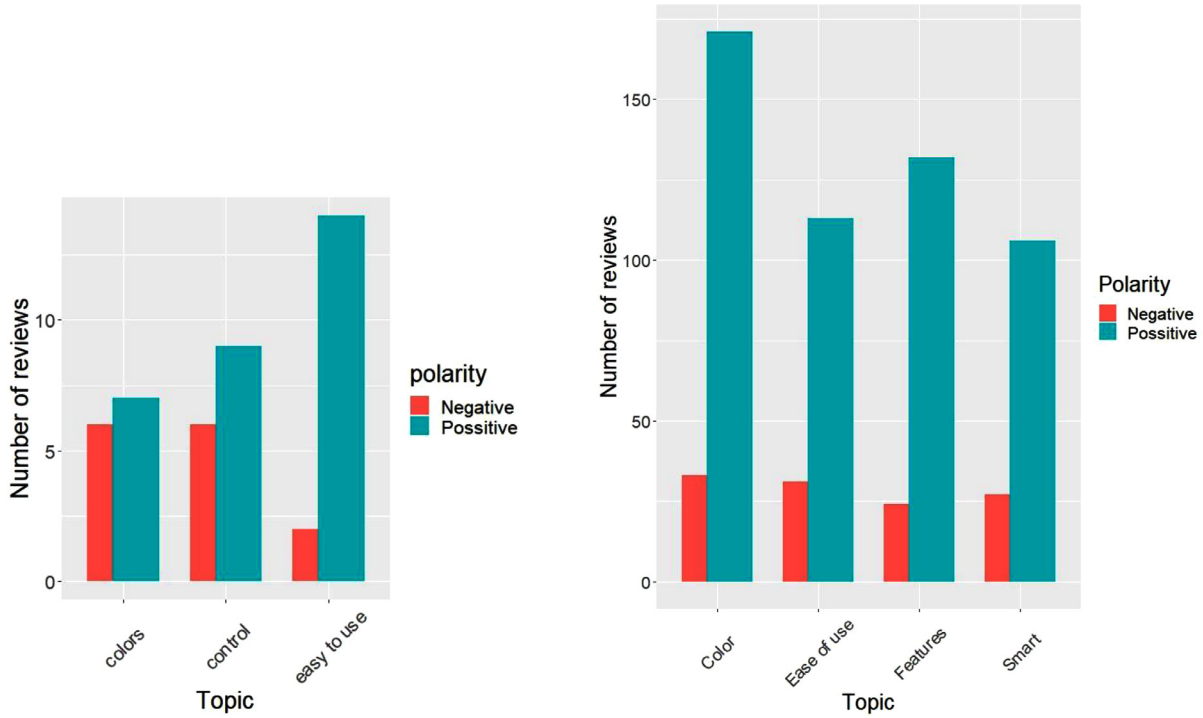



**Fig. 10.** Number of occurrence in peer reviewed articles on yearly basis.




(a) Sengled Smart Light Bulbs - color changing light bulbs

(b) Philips Hue White and Color Ambiance A19 60W Equivalent LED Smart Bulb Starter Kit

**Fig. 11.** Aspect based sentiment analysis: Smart lighting products.

*of pee (warm white)*" another example is "*.. I love being able to set different themes in the living room depending on what I am doing or the time of day.*".

*ease of use* had noticeable attention among users. For instance "*these bulbs are awesome! Very easy to set up and get operational.*" and "*..I found the set up easy, the pairing worked effortlessly,*".

The ability to control the lights and colors with app, by voice commands through Amazon alexa and Google Home gave user a sense of power. As a result, the reviews mostly included positive feedback with little negative feedback. For instance "*..Combining this with a tap switch, Alexa, and the Philips Hue app makes this the killer smart light setup. "Alexa, make my living room magenta". "Alexa, dim the kitchen by 13%". And it just works."* while another one commented that "*..issue with all the 'sync your music to your lights' apps for sure - not quite snappy enough to make that a thing, yet.*". It seems that while some users were enjoying the control and features, it was not the case for all the users.

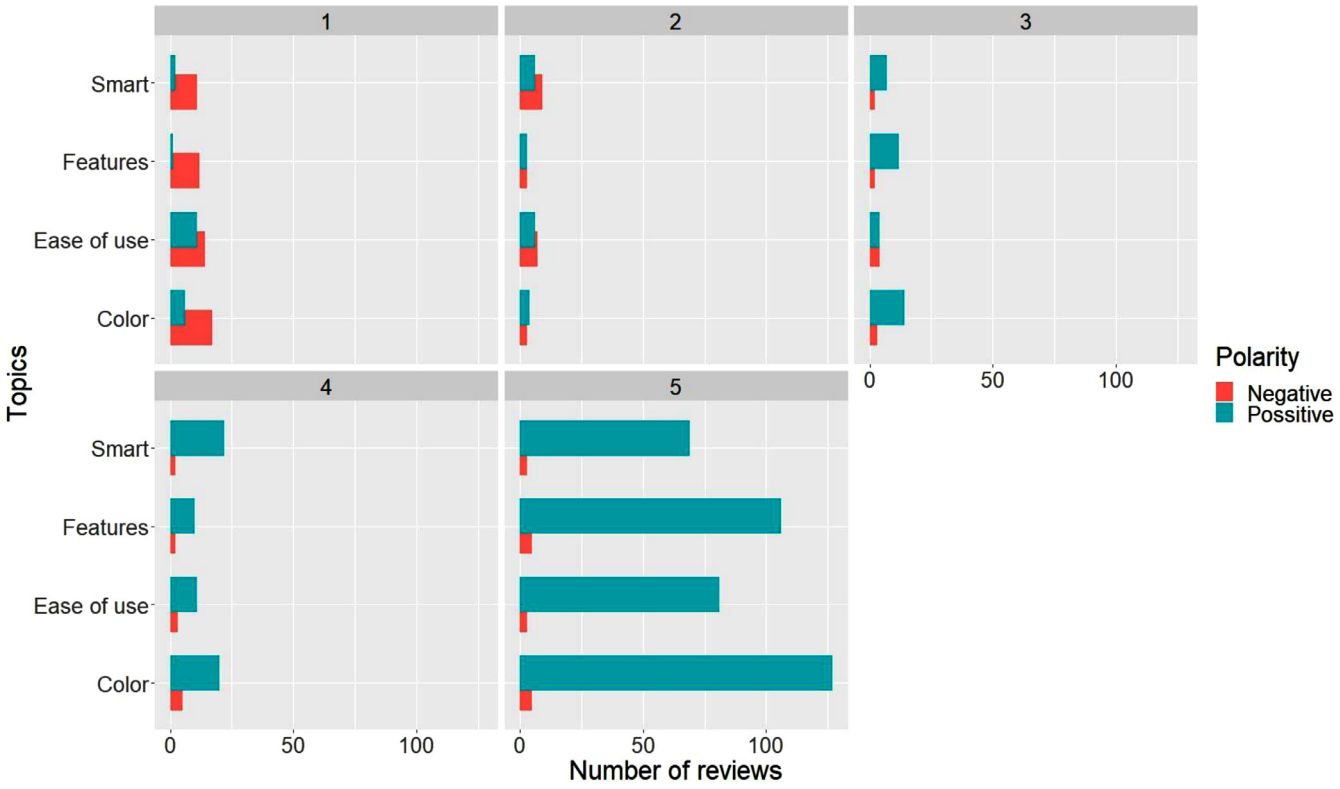



**Fig. 12.** Aspect based sentiment analysis in two extremes.

### 3.3. Star-aspect based sentiment analysis

We analyzed the Hue starter kit sentiments, based on two different components. One, the star specified by user at the time of review, and two, the time of the review. The product star is a number that ranges between 1 and 5, 1 being the lowest, and 5 being the highest. We analyzed the star product to distinguish the satisfied and unsatisfied users with regards to each separate topic. Therefore, the concerns related to each topic can be highlighted and addressed. Fig. 12 shows the number of positive and negative reviews with regard to each aspect, and in each star category. Although, overall, users were satisfied with the product, there were some negative comments about each aspects of the product. For instance, *features and control*; majority of customers were satisfied with the features and control ability of the product in star 5 category. However, in star 1 category some users were not satisfied with features and control ability of the product. This may suggest that users' preferences with regards to control and automation are diverse. Therefore, lighting business can provide solutions that customize and tailor the control and features option according to each user's needs. The result of aspect based sentiment analysis based on the time of review shows how users' sentiments changed throughout the time, as shown in Fig. 13. This product was introduced in August 2017 and therefore high number of reviews in 2018 is a result of maturity phase that is being followed by reduction of interest in 2019 and 2020. Prior research also confirms that users' excitement and engagement decreases overtime. As a result, in order to keep users' interest, it may be beneficial to periodically change the features and controls in lighting system.

## 4. Discussion

In this research we proposed a framework to utilized two publicly available data sources to bridge possible gaps between research and development and provide research agenda in smart lighting business. Lighting is a major electricity consumer in buildings (Ryckaert et al., 2010) and even an small improvement, if adopted wildly, can have a huge impact. Smart lighting devices have high potential for energy savings. However, the adoption rate of this newly developed system is low for average users (Shin et al., 2018). In this study, we analyzed the perception of the user as well as topics and trends in literature concerning smart lighting business. The contribution of proposed framework is discussed in terms of literature and practical application in following two sections:

### 4.1. Contribution to literature

There have been some attempts to identify gaps between research and practice in the research community, for instance, Zou, Sunindijo, & Dainty (2014) reviewed articles in construction safety in terms of their relationship with safety knowledge, safety learning processes, and safety management practices. They acknowledged the gap between research and the practical needs of the industry in construction safety and advocated for the use of qualitative and quantitative research methods to help integrate the realms of theory and practice. There are multiple research institutions around the world and some of these research projects are published and therefore, are available for public review.

In today's world, online shopping is a great source for purchasing household needs that enables more informed decisions. In an online purchasing platform, products can be compared with one another in terms of usage, cost etc. Most of online shopping platforms have review sections in which customers can provide feedback, and review other consumers' feedback. Leveraging this valuable source of information requires an automated and efficient framework. In this study, we consider available journal articles' abstract and published reviews on Amazon as research data. After a product is developed and released for public use, user perceptions define how a product is adopted and guide later development paths.

Researchers are advancing knowledge and easing our everyday life, and publishing result of their work. Reviewing and analyzing a vast

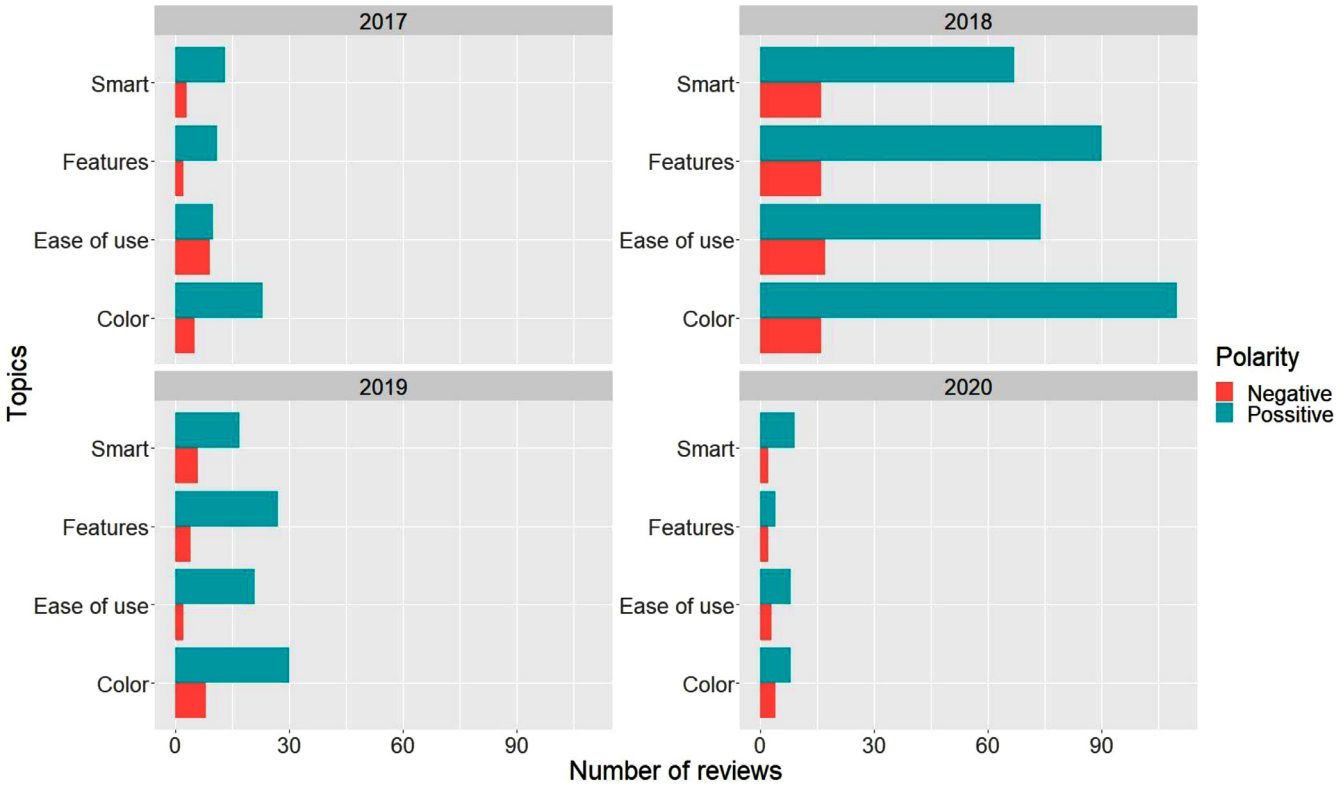



**Fig. 13.** Aspect based sentiment analysis through out the time.

amount of knowledge in literature requires intensive human work, and efficient analysis of literature requires AI and more advanced technologies. Analyzing the trend and topics in literature, finding gaps, and proposing future directions are of a great interest among researchers. In today's world, reviewing literature and summarizing and analyzing trends can be done efficiently with the help of AI. Topic modeling has been implemented for this purpose in several studies (Abuhay, Nigatie, & Kovalchuk, 2018; Älgå, Eriksson, & Nordberg, 2020; Chen, Wang, Pan, & Xiong, 2019; Kim, Park, & Lee, 2020; Kumar, Kar, & Ilavarasan, 2021; Kushwaha, Kar, & Dwivedi, 2021; Sun & Yin, 2017; Wang et al., 2018; Zou, 2018).

In this study we utilize text mining for analyzing high amount of unstructured data (online reviews and journal articles). In this line of research text mining has been utilized for text analysis application such as sentiment analysis (Neogi, Garg, Mishra, & Dwivedi, 2021); however, Bag Of Word (BOW) and term frequency are among frequently used methods to summarize a text document. In this work we propose a probabilistic aspect based sentiment analysis framework to analyze users perception. We identify aspects and analyze the sentiment toward those topics. Studies utilized LDA for topic modeling (Adikari, Burnett, Sedera, de Silva, & Alahakoon, 2021; Karami, Bookstaver, Nolan, & Bozorgi, 2021; Rajendran & Sundarraj, 2021) and use vector of words to represent a document. However, in this study we use bigrams as phrases and utilize LDA to generate informative topics. This approach is different that traditional vector space modeling, where BOW is input. It has been found that phrase based topic modeling has significant improvement, specially on short text data (Kherwa & Bansal, 2020; Nokel & Loukachevitch, 2016). Our trend analysis suggested that smartness, users comfort, and energy consumption have gained a lot of attention in recent years, and as expected, energy was the most attractive topic among researchers.

Summary of contributions of this work to literature are as follow: 1- it contributes to our understanding of possible gaps between the research and development, particularly in smart lighting business. Results of this analysis showed that while energy consumption is a very important subject among research community, the end user is not considering that as a factor in their discussion and reviews. 2- it creates a path toward further analysis in this area, and as a result, helps in bridging possible gaps. This framework enables analysis of big, unstructured text data sources to generate insights to help the research and development vision. 3 - it provides insights that help in analyzing users' perceptions and needs. Analysing end users' concern through topic modeling and sentiment analysis provides input for research and product development in smart lighting business and can help with the adoption of these newly introduced products.

### 4.2. *Implications for practice*

Smart market is in the early stage of its development. Understanding users perception, understanding research in this line of work and aligning those into same direction is of a great importance for smart home products. This study proposes an automated framework to leverage user generated data, and research articles in an efficient manner to understand these two value able sources of information. This framework can be utilized for research and development purposes for smart home products. Moreover, with the development of smart technologies, new characteristics and concerns arise that would need to be explored and this framework can be used for this purpose. This work can be used to understand the users perception and align those with research line. Aligning research with development considering end user's perception would increase the adoption rate of smart lighting system, and as a result, improve the energy efficiency of buildings. It has been shown that strategies to market the usefulness of a smart product would increase the smart home market (Shin et al., 2018). Moreover, emotional value and impression of a product, significantly contributes to the brand's favor (Kato, 2021).

This study shows that users are excited about color changing ability of smart lighting products and they seem to not consider electricity con-

sumption as a factor in describing the usefulness of smart lighting products. Similarly, electricity consumption was not a concern among other smart home product users such as smart thermostat (Koupaei, Song, Cetin, & Im, 2020). Understanding a product's utility, directly from customers is a major factor that affects both the adoption and purchasing time of smart homes services (Shin et al., 2018). Therefore, proposing and promoting smart features such as color changing ability, by leveraging IoT enabled data and AI can impact the recognition of usefulness and impact adoption rate. Such innovative ideas are to be explored by capitalizing IoT enabled data and AI, for instance, generating personalized recommendations in light usage. A study showed that older people with higher income are the main target consumers for smart homes; therefore, it is crucial to understand how to make these products intuitive and more convenient for older age groups, unfamiliar with new technologies (Shin et al., 2018). Our study confirms that ease of use is indeed an important factor among people who purchased the product. Analyzing users behavior (for instance wake up and go to sleep patterns) can be another another interesting and innovative smart feature that can ease the use of smart products. The result of this analysis suggests that smartness is one of the factors that excites the users. Proposed insights suggest that in order to increase the adoption rate of smart lighting products, the focus of marketing could be the promotion of smartness features and ease of use, rather than sustainability. This knowledge is critical to increase the adoption rate of smart lighting products, and as result of mass adaptation, significantly increase the energy efficiency.

## 5. Conclusion

In this study, we attempted to analyze possible gaps and direction in research community with development and provide research agenda. For this purpose, we utilized two publicly available data sources. We considered peer reviewed journal articles to analyze research and online reviews related to two smart lighting products for development. We utilized text mining and aspect based sentiment analysis, to revel possible gaps between research and development. We analyzed and compared the topics discussed in these two sources of data in terms of direction, date, and polarity.

Results revealed that the energy consumption and related infrastructure is the major concern of the research community, while, the users seemed more interested in the features of the products such as color changing ability. Smartness is one of the topics that attracted attention of both the research community and end users in smart lighting products. Aspect based sentiment analysis reveled that users were excited about smart lighting products in general and the highest attention was toward *color* aspect of these products. Analyzing the sentiments in each category of stars (1–5) showed that people have diverse preferences. Moreover, analyzing the time of reviews, suggested that each product has a maturity phase, and afterwards, users engagement with the system decreases. Therefore, in order to keep users interested in such products, it would be beneficial to provide customized and personalized features such as (routine, color,.) recommendations ()(Zarindast & Wood, 2021) in line of making smart products. On the other hand, as it seems that users are not much concerned about energy consumption, providing smart products with energy efficient control decisions by default, meanwhile attracting users attention with other novel advantages such as *color changing* ability can be a good alternative.

Overall, result of this work has major impact toward understanding possible research and development gaps, and can potentially help bridging possible gaps in smart lighting business. Understanding users' perception may be beneficial for research and development to increase the adjustment rates and provide desired solutions. It is notable that reviews are filtered based on language and are provided in the US. Future work can benefit from analyzing a broader range of users and including potential users. It may also be beneficial to develop field specific database for sentiment analysis.

## Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, ornot-for-profit sectors.

## References

Abuhay, T. M., Nigatie, Y. G., & Kovalchuk, S. V. (2018). Towards predicting trend of scientific research topics using topic modeling. *Procedia Computer Science, 136*, 304–310.
Adikari, A., Burnett, D., Sedera, D., de Silva, D., & Alahakoon, D. (2021). Value co-creation for open innovation: An evidence-based study of the data driven paradigm of social media using machine learning. *International Journal of Information Management Data Insights, 1*(2), 100022.
Al Dakheel, J., Del Pero, C., Aste, N., & Leonforte, F. (2020). Smart buildings features and key performance indicators: A review. *Sustainable Cities and Society, 61*, 102328.
Alam, F., Ofli, F., & Imran, M. (2020). Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes Harvey, Irma, and Maria. *Behaviour & Information Technology, 39*(3), 288–318.
Älgå, A., Eriksson, O., & Nordberg, M. (2020). Analysis of scientific publications during the early phase of the COVID-19 pandemic: Topic modeling study. *Journal of medical Internet Research, 22*(11), e21559.
Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications (IJACSA), 6*(1).
Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on big data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics, 24*(1), 1–7.
Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.
Building Energy Data Book (2011). Us department of energy, office of energy efficiency and renewable energy, 2005. Available: http://buildingsdatabook.eren.doe.gov,.
Caicedo, D., Pandharipande, A., & Leus, G. (2011). Occupancy-based illumination control of LED lighting systems. *Lighting Research & Technology, 43*(2), 217–234.
Chatterjee, S., Kar, A. K., & Gupta, M. P. (2018). Success of IoT in smart cities of India: An empirical analysis. *Government Information Quarterly, 35*(3), 349–361.
Chen, H., Wang, X., Pan, S., & Xiong, F. (2019). Identify topic relations in scientific literature using topic modeling. *IEEE Transactions on Engineering Management*.
Chew, I., Karunatilaka, D., Tan, C. P., & Kalavally, V. (2017). Smart lighting: The way forward? Reviewing the past to shape the future. *Energy and Buildings, 149*, 180–191.
Choi, H. S., Lee, W. S., & Sohn, S. Y. (2017). Analyzing research trends in personal information privacy using topic modeling. *Computers & Security, 67*, 244–253.
DOE Buildings Energy Data Book (2009). Department of energy, march 2009.
Gethers, M., & Poshyvanyk, D. (2010). Using relational topic models to capture coupling among classes in object-oriented software systems. In *2010 IEEE international conference on software maintenance* (pp. 1–10). IEEE.
Greene, D., & Cross, J. P. (2015). Unveiling the political agenda of the european parliament plenary: A topical analysis. In *Proceedings of the ACM web science conference* (pp. 1–10).
Gunay, H. B., O'Brien, W., Beausoleil-Morrison, I., & Gilani, S. (2017). Development and implementation of an adaptive lighting and blinds control algorithm. *Building and Environment, 113*, 185–199.
Harkare, A., Potdar, V., Mishra, A., Kekre, A., & Harkare, H. (2021). Methodology for implementation of building management system using IoT. In *Evolutionary computing and mobile sustainable networks* (pp. 939–948). Springer.
Heydarian, A., Carneiro, J. P., Gerber, D., & Becerik-Gerber, B. (2015). Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance. *Building and Environment, 89*, 217–228.
Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. In *Cognitive informatics and soft computing* (pp. 639–647). Springer.
Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169–15211.
Jung, H., & Lee, B. G. (2020). Research trends in text mining: Semantic network and main path analysis of selected journals. *Expert Systems With Applications, 162*, 113851.
Karami, A., Bookstaver, B., Nolan, M., & Bozorgi, P. (2021). Investigating diseases and chemicals in COVID-19 literature with text mining. *International Journal of Information Management Data Insights, 1*(2), 100016.
Kato, T. (2021). Functional value vs emotional value: A comparative study of the values that contribute to a preference for a corporate brand. *International Journal of Information Management Data Insights, 1*(2), 100024.
Kherwa, P., & Bansal, P. (2020). Semantic N-gram topic modeling. *EAI Endorsed Transactions on Scalable Information Systems, 7*(26).
Kim, S., Park, H., & Lee, J. (2020). Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis. *Expert Systems with Applications, 152*, 113401.
Knez, I., & Kers, C. (2000). Effects of indoor lighting, gender, and age on mood and cognitive performance. *Environment and Behavior, 32*(6), 817–831.
de Korte, E. M., Spiekman, M., Hoes-van Oeffelen, L., van der Zande, B., Vissenberg, G., Huiskes, G., & Kuijt-Evers, L. F. M. (2015). Personal environmental control: Effects of pre-set conditions for heating and lighting on personal settings, task performance and comfort experience. *Building and Environment, 86*, 166–176.
Koupaei, D. M., Song, T., Cetin, K. S., & Im, J. (2020). An assessment of opinions and perceptions of smart thermostats using aspect-based sentiment analysis of online reviews. *Building and Environment, 170*, 106603.

Krallinger, M., & Valencia, A. (2005). Text-mining and information-retrieval services for molecular biology. *Genome Biology, 6*(7), 224.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(1), 100008.

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), 100017.

Le, D. N., Le Tuan, L., & Tuan, M. N. D. (2019). Smart-building management system: An internet-of-things (IoT) application business model in Vietnam. *Technological Forecasting and Social Change, 141*, 22–35.

Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., & Anzanello, M. J. (2020). Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management, 83*, 101760.

Minoli, D., Sohraby, K., & Occhiogrosso, B. (2017). IoT considerations, requirements, and architectures for smart buildings-energy optimization and next-generation building management systems. *IEEE Internet of Things Journal, 4*(1), 269–283.

Mishra, A., Irwin, D., Shenoy, P., Kurose, J., & Zhu, T. (2013). GreenCharge: managing renewable energy in smart buildings. *IEEE Journal on Selected Areas in Communications, 31*(7), 1281–1293.

Moadab, N. H., Olsson, T., Fischl, G., & Aries, M. (2021). Smart versus conventional lighting in apartments-electric lighting energy consumption simulation for three different households. *Energy and Buildings, 244*, 111009.

Mortenson, M. J., & Vidgen, R. (2016). A computational literature review of the technology acceptance model. *International Journal of Information Management, 36*(6), 1248–1259.

Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights, 1*(2), 100019.

Nicol, F., Wilson, M., & Chiancarella, C. (2006). Using field measurements of desktop illuminance in European offices to investigate its dependence on outdoor conditions and its effect on occupant satisfaction, and the use of lights and blinds. *Energy and Buildings, 38*(7), 802–813.

Nokel, M., & Loukachevitch, N. (2016). Accounting ngrams and multi-word terms can improve topic models. In *Proceedings of the 12th workshop on multiword expressions* (pp. 44–49).

Peruffo, A., Pandharipande, A., Caicedo, D., & Schenato, L. (2015). Lighting control with distributed wireless sensing and actuation for daylight and occupancy adaptation. *Energy and Buildings, 97*, 13–20.

Porturas, T., & Taylor, R. A. (2020). Forty years of emergency medicine research: Uncovering research themes and trends through topic modeling. *The American Journal of Emergency Medicine*.

Rajendran, D. P. D., & Sundarraj, R. P. (2021). Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings. *International Journal of Information Management Data Insights, 1*(2), 100027.

Ryckaert, W. R., Lootens, C., Geldof, J., & Hanselaer, P. (2010). Criteria for energy efficient lighting in buildings. *Energy and Buildings, 42*(3), 341–347.

Shamsuzzoha, A., Niemi, J., Piya, S., & Rutledge, K. (2021). Smart city for sustainable environment: A comparison of participatory strategies from helsinki, singapore and london. *Cities, 114*, 103194.

Shin, J., Park, Y., & Lee, D. (2018). Who will be smart home users? An analysis of adoption and diffusion of smart homes. *Technological Forecasting and Social Change, 134*, 246–253.

Sievert, C., & Shirley, K. (2014). LDAvis: a method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63–70).

Sun, L., & Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C: Emerging Technologies, 77*, 49–66.

Wagiman, K. R., Abdullah, M. N., Hassan, M. Y., & Radzi, N. H. M. (2021). A new metric for optimal visual comfort and energy efficiency of building lighting system considering daylight using multi-objective particle swarm optimization. *Journal of Building Engineering, 43*, 102525.

Wang, L., Lakin, J., Riley, C., Korach, Z., Frain, L. N., & Zhou, L. (2018). Disease trajectories and end-of-life care for dementias: latent topic modeling and trend analysis using clinical notes. In *AMIA annual symposium proceedings: vol. 2018* (p. 1056). American Medical Informatics Association.

Xu, L., Pan, Y., Yao, Y., Cai, D., Huang, Z., & Linder, N. (2017). Lighting energy efficiency in offices under different control strategies. *Energy and Buildings, 138*, 127–139.

Yun, E. (2020). Review of trends in physics education research using topic modeling. *Journal of Baltic Science Education, 19*(3), 388–400.

Yun, J., & Geum, Y. (2020). Automated classification of patents: A topic modeling approach. *Computers & Industrial Engineering, 147*, 106636.

Yun, Y., Hooshyar, D., Jo, J., & Lim, H. (2018). Developing a hybrid collaborative filtering recommendation system with opinion mining on purchase review. *Journal of Information Science, 44*(3), 331–344.

Zarindast, A., Wood, J., & Sharma, A. (2021) A data-driven personalized smart lighting recommender system. arXiv preprint arXiv:2104.02164

Zarindast, Atousa, & Wood, Jonathan (2021). A data driven personalized smart lighting recommender system. *Frontiers in big data -data mining and management* Submitted for publication.

Zou, C. (2018). Analyzing research trends on drug safety using topic modeling. *Expert Opinion on Drug Safety, 17*(6), 629–636.

Zou, P. X. W., Sunindijo, R. Y., & Dainty, A. R. J. (2014). A mixed methods research design for bridging the gap between research and practice in construction safety. *Safety Science, 70*, 316–326.