ANALYZING PHYSICAL WORKPLACE AND SERVICE MANAGEMENT USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING APPROACHES

A Dissertation
Presented to
The Academic Faculty

by

Sungil Hong

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Building Construction

Georgia Institute of Technology
May 2022

ANALYZING PHYSICAL WORKPLACE AND SERVICE MANAGEMENT USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING APPROACHES

Approved by:

Dr. Eunhwa Yang, Advisor
School of Building Construction
Georgia Institute of Technology

Dr. Daniel Castro-Lacouture
School of Building Construction
Georgia Institute of Technology

Dr. Baabak Ashuri
School of Building Construction/ Civil & Environmental Engineering
Georgia Institute of Technology

M.Sc. Bonnie Sanborn
DLR Group

Dr. Xiuwei Zhang
School of Computational Science and Engineering
Georgia Institute of Technology

Date Approved: April 13, 2022

For Inyoung and Leah

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Professor Eunhwa Yang, who is a thoughtful and patient educator, for her invaluable supervision and support during the course of my PhD degree. I have been inspired by her passion for research on human and built environments. Also, Dr. Baabak Ashuri has helped me refine the quality of my research. His acute insight has pushed me to raise the bar, further my limits, and produce fruitful outcomes. Dr. Xiuwei Zhang graciously has become my committee member notwithstanding my major. She has given me concrete guidance in computational approaches, which allowed me to be confident in my research methodology. Dr. Daniel Castro has been a huge supporter throughout my Ph.D. studies. His love for students and steadfast support have made this challenging journey steady and enjoyable. Bonnie Sanborn has provided me with practical industry views and embodied the wonderful compatibility of a practitioner and researcher. I have gained the equilibrium to carry on with my career after earning my degree. I am grateful to have been a student of all of my committee members.

I would also like to thank my friends and colleagues who have rooted for me to complete my degree and have not hesitated to help me. I feel privileged to have such a great support system. My journey was enjoyable and meaningful because of my friends: Soobum Kim, Yujin Kim, Junghyun Kim, Jung Hyun Lee, Seunghun Jwa, Yun Joon Jung, Frederick Chung, Sungjin Kim, Hoon Na, Puting Yo, Gemma Fiduk, Catherine Bisson, Tom Grice, Ran Zhang, Yuming Zhou, Austin Slater, and Mina Park. This single page is not enough to express my appreciation for you, my dear friends. I will try to be a good friend to you and help you throughout our lives.

My gratitude, of course, extends to my family. My wife, Inyoung, you are home in my heart. You have always loved and trusted me whatever I do and wherever I am. I could not have completed this long journey without your trust and love. I am grateful and proud to be your husband. Now, it is my turn to support you to achieve your dream. My precious daughter, Leah, your birth has meant so much to me. You opened my eyes and heart to live with hope. Inyoung and I look forward to the future with you. I would also like to thank my parents and parents-in-law who have always prayed for me with unconditional love and unwavering support. To my older sister and brother-in-law, I have learned how to become a breadwinner from you. You are an inspiration in my life. To my wife's brother and sister-in-law, I have learned how true leaders behave. I appreciate your thoughtfulness and care for my family. I love you with my whole heart.

Last but not least, thank you, God, for sending me to the United States and protecting my family anywhere and anytime. It has been a time to know more of you. I do not know how my life will go afterward, but I believe that you know better than I. One thing I do know is that I would like to live and love my neighbors as you have loved me. *"Love one another. As I have loved you, so you must love one another."* [John 13:34], *"Be kind and compassionate to one another, forgiving each other, just as in Christ God forgave you."* [Ephesians 4:32]

TABLE OF CONTENTS

ACKNOWLEDGEMENTS **iv**

LIST OF TABLES **x**

LIST OF FIGURES **xi**

LIST OF SYMBOLS AND ABBREVIATIONS **xii**

SUMMARY **xiii**

CHAPTER 1. Introduction **1**

1.1 Workplace Demand Changes **1**

1.2 Redefined Facility Management (FM) Model **3**

CHAPTER 2. A Review of Physical Environment and Service In Workplace **7**

2.1 Physical Work Environment **7**

2.1.1 Ambient Environment 7

2.1.2 Office Layout 8

2.1.3 Outside View 9

2.1.4 Design Aesthetics 9

2.1.5 Technology and Facilities 10

2.2 Service in Workplace **11**

2.2.1 Administrative Support 11

2.2.2 Workplace Education 12

2.2.3 Health Promoting Programs 12

CHAPTER 3. Physical Environment and Service In Coworking Space 13

3.1 Physical Work Environment in Coworking Space 15

3.1.1 Ambient Environment 15

3.1.2 Space Layout 16

3.1.3 Interior and Aesthetics 17

3.1.4 Technology and Facilities 17

3.2 Service in Coworking Space 18

3.2.1 Administrative Support 18

3.2.2 Workplace Education 18

3.2.3 Socialization and Networking Service 19

3.2.4 Leisure and Well-being 20

3.3 Challenges in Coworking Space Research 21

3.4 Overarching Research Objectives 23

CHAPTER 4. Study I: Thematic categorization and user preference analysis of coworking spaces during COVID-19 using social media data 26

4.1 Introduction 26

4.2 Methods 32

4.2.1 Data Collection and Cleaning 35

4.2.2 Mixed-methods Approach: Content Analysis, Classification, and Clustering 36

4.2.3 Model Validation 41

4.3 Results 42

4.3.1 Content Analysis with Sentiment Analysis 42

4.3.2 Descriptive Analysis 48

4.3.3 Machine Learning: Clustering and Topic Discovery 50

4.3.4 Machine Learning: Multi-label Text Classification 52

4.4 Discussion 55

4.4.1 Dissatisfaction with Physical Environment and Service Categories in CWS 55

4.4.2 Changes in User Preferences for CWS Factors During the Pandemic 57

4.4.3 Insight from Clustering Outcomes 61

4.4.4 Model Validation of Classifiers 62

4.5 Limitations 63

CHAPTER 5. Study II: Data Integration to Build Robust Machine Learning Based Text Classification Models for Facility Maintenance Data 65

5.1 Introduction 65

5.2 Methods 69

5.2.1 Framework 69

5.2.2 Data Collection and Cleaning 70

5.2.3 Category Matches and Feature Engineering 71

5.2.4 Processing through NLP 75

5.2.5 Machine Learning based Multi-class Classification Models 77

5.2.6 Evaluation 79

5.3 Results 81

5.3.1 Feature Engineering 81

5.3.2 Model Validation 83

5.4 Discussion 87

5.4.1 Feature Engineering and Adaptability 88

5.4.2 Improved Classification Performance with Difference Datasets 90

5.4.3 Implications 92

5.5 Limitations 93

CHAPTER 6. Discussion 94

6.1 User Preferences of Facilities and Services during COVID-19 94

6.2 Applications of NLP and Machine Learning Techniques 97

6.3 Validation by Professionals 99

CHAPTER 7. Conclusion 102

APPENDIX A. Appendices 104

A.1 Study I: Category Labeling Schemes 104

A.2 Study I: Hyper-parameters 113

A.3 Study II: Confusion matrices 114

REFERENCES 116

LIST OF TABLES

Table 1. Thematic categories of coworking space 34

Table 2. Overall user experiences in CWS during COVID-19 (N=455) 43

Table 3. COVID-19 related user experiences in CWS 47

Table 4. Cooccurring category rankings before and during COVID-19 50

Table 5. Average overall accuracy and micro average f1 score of predictions or each classifier about the test data set 53

Table 6. MEP maintenance categories and sub-categories 74

Table 7. Confusion matrix 80

Table 8. Chi-square feature selection of maintenance requests 82

Table 9. Cross-validation results of the four classifiers (campus building) 84

Table 10. Comparison of accuracy/ f1 score between the ATDC and the combined datasets 85

Table 11. CWS management strategies during COVID-19 96

Table 12. Summary of the applications of NLP and machine learning techniques 99

LIST OF FIGURES

Figure 1. Conventional FM model: services supporting occupants and building owners .. 4

Figure 2. Redefined FM model: FM as a space & service provider 6

Figure 3. Overview of studies to build CWS management strategies 25

Figure 4. The flow of text mining and analysis 33

Figure 5. Locations of the Yelp reviews collected 35

Figure 6. Sentiment analysis example using VADER 37

Figure 7. Multi-label example of a CWS Yelp review 38

Figure 8. The proportion of CWS reviews about before and during COVID-19 49

Figure 9. Reorganized clustering results based on the thematic categories...................... 52

Figure 10. (a) Category level classification performance of prediction accuracy (left), and (b) F1 score (right) 54

Figure 11. Relationship diagram of CWS experience 60

Figure 12. Data integration of different maintenance datasets 70

Figure 13. Illustration of pre-processing... 75

Figure 14. Confusion matrices of the text classifications in HVAC................................. 86

Figure 15. Confusion matrices of the text classifications in electrical 114

Figure 16. Confusion matrices of the text classifications in plumbing........................... 115

LIST OF SYMBOLS AND ABBREVIATIONS

BERT	Bidirectional Encoder Representations from Transformers
CMMS	Computerized Maintenance Management System
CWS	Coworking Space
EWOM	E-word of mouth (EWOM)
FM	Facility Management
FP	False Positives (FP)
HVAC	Heating, Ventilation, Air-conditioning
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications
IAQ	Indoor Air Quality
IEQ	Indoor Environmental Quality
LR	Logistic Regression
MLSMOTE	Multi-label Synthetic Minority Over-sampling Technique
MLP	Multi-layer Perceptron (MLP)
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
RBF	Radial Basis Function
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
FN	False Negatives
TN	True Negative
TP	True Positive
UMAP	Uniform Manifold Approximation and Projection for dimension reduction
VADER	Valence Aware Dictionary for Sentiment Reasoning

SUMMARY

The demand for workplace flexibility has emerged according to ever-changing environments, such as sharing and gig economy, alternative work arrangement, and COVID-19. This study proposes a redefined facility management model corresponding to the changing circumstances, which provides not only space but also activity support and leisure services. Coworking space (CWS) is one of the embodiments of the model. This research aims to develop CWS management strategies for 1) user preferences in physical workplace environments and services during COVID-19 and 2) data management methods utilizing natural language processing (NLP) and machine learning techniques. Two main studies in this research address three research objectives: 1) identifying preferences for facilities and services factors in CWSs during COVID-19; 2) detecting changing preferences for factors about facilities and services during COVID-19; 3) proposing the applications of machine learning and NLP techniques and demonstrating the applicability of computational data collection and analysis methods in the physical workplace management research. First, Study I proposes a thematic categorization scheme of CWS spatial and service factors and elements. Based on the categories, a mixed-method approach was utilized for the comprehensive data analysis, including content analysis, classification, and clustering. The results show that CWS users have become sensitive to disruptive behaviors and hygienic responses to infectious diseases after the pandemic. The findings also present a need for a sense of community and various technology needs for virtual interactions. Second, Study II performed the data integration of a large computerized maintenance management system dataset of a public college campus into a single CWS building maintenance dataset to build robust machine learning-based text classification models for a small dataset. The results show the qualitative and quantitative increase in prediction performance of text classifications. Study II implies that data integration will accelerate smart facility management, including small or single buildings, by sharing public datasets. In conclusion, this research sheds light on online big data collection and analysis in physical workplace management research. It also presents how the facility management industry can apply such state-of-the-art technology in utilizing historical data to make data-driven decisions.

CHAPTER 1. INTRODUCTION

1.1 Workplace Demand Changes

Several changes in external environmental factors have a significant impact on the workplace, namely workforce demographics, technology advancements, and pandemics. Millennials and Generation Z will make up, if they have not already, the majority of the American workforce, with most recent official census figures showing they account for 49.1% of the U.S. working population in 2017 (United States Census Bureau 2018). These generations have distinct characteristics from older generations. First, millennials and Generation Z want to be assured of independence in their work, strongly preferring to work independently (Petriglieri, Ashford, and Wrzesniewski 2018). Also, they prioritize opportunities for advancement and having meaningful experiences at work (Caraher 2016; Ng, Schweitzer, and Lyons 2010). These attributes have led to the rise of the gig economy where they can work on a project basis.

Besides the working demographic shifts, technology advancements affect work environments. High-performance computing power has accelerated the use of artificial intelligence (AI) technology which can eventually replace task-level jobs and routine work (Huang and Rust 2018). In addition, advanced information communication technologies (ICTs) with mobile devices and operating systems facilitate the emergence of diverse digital platforms (de Reuver, Sørensen, and Basole 2018) as well as enhance the experience of virtual meetings and collaboration online (DeFilippis et al. 2020). Specifically, digital platforms and online meeting tools promote “sharing and collaborative exchanges” as a mediating technology (Sutherland and Jarrahi 2018).

Due to the changes in the working population and technology advancements, the sharing economy has thrived. Millennials and Generation Z define themselves as travelers not bound to an assigned place and who eschew material possessions such as a car and house except for digital nomad possessions (Goldman Sachs 2019). The sharing economy ranges from service exchanges to productive assets with interactions mainly taking place through a mobile platform (Schor 2016). The sharing economy is in synergy with technology advance in digital platforms and AI that accelerate knowledge work and the gig economy and support alternative work arrangements (AWA) (Alpaydin 2020; Huang and Rust 2018) because knowledge workers are flexible in employment, schedules, and places (Spreitzer, Cameron, and Garrett 2017).

Besides these two external factors, COVID-19 critically changed work patterns. The pandemic forced knowledge workers to work from home. The percentage of people working from home in the U.S. increased from 8.2% in February 2020 to 35.2% in May 2020 (Bick, Blandin, and Mertens 2020). Working from home made work hours more flexible and longer (Singer-Velush, Sherman, and Anderson 2020). Commuting time was repurposed. Instead of formal and informal face-to-face meetings, planned meetings online increased by 10%. Most communications have occurred through virtual meetings during the COVID-19 pandemic (DeFilippis et al. 2020). A prolonged period of enforced working from home has changed the perception of full-time work in offices. For instance, working from home has stimulated employees' preference for working from home because of the increased flexibility which allows for more work and life balance, and reduced commuting times (Boland et al. 2020). The observed environmental factors led to changes in workplace demands such as requiring workplace flexibility and resiliency.

The workplace is 'the physical settings in which work happens, to the services that support people in those settings and, perhaps most critically, the management process that enable their effective use' (Alexander et al. 2004, p. 4). Thus, the changes in workplace demands require changes in space and service in the workplace. Since satisfaction with one's work environment is associated with overall job satisfaction and retention as well as individual and organizational performance (Carlopio 1996; Jain and Kaur 2014; Lee and Brand 2005), the changing demands for the workplace should be addressed in a deliberate manner.

Coworking spaces (CWSs) are one of the alternative workplaces that have adapted to the aforementioned environmental factors and workplace attributes. Since COVID-19, to accomplish work flexibility and suburbanization, CWSs are attracting not only startup and small-to-medium (SME) employees, but also traditional office workers wishing to work closer to home for better performance (Smith et al. 2020). This trend leads to the necessity to investigate the properties of CWSs from both spatial and service aspects that can accommodate different types of knowledge workers.

1.2 Redefined Facility Management (FM) Model

In parallel with the attributes of CWSs, this research is interested in a new facility management (FM) model that reacts to the aforementioned environmental factors. Conventionally, FM service providers have addressed both hard services such as building maintenance and janitorial services and soft issues such as human resources (HR), safety, and security of the relevant organizations, namely building owners and tenants, to support their core activities (Atkin and Brooks 2015). FM strategies have focused on how to support the core business of these organizations in the effective utilization of infrastructure and facilities while reducing costs (Alexander et al. 2004). A FM company is in indirect relation to designing and using the spaces as an adviser or a supporter shown in Figure 1.

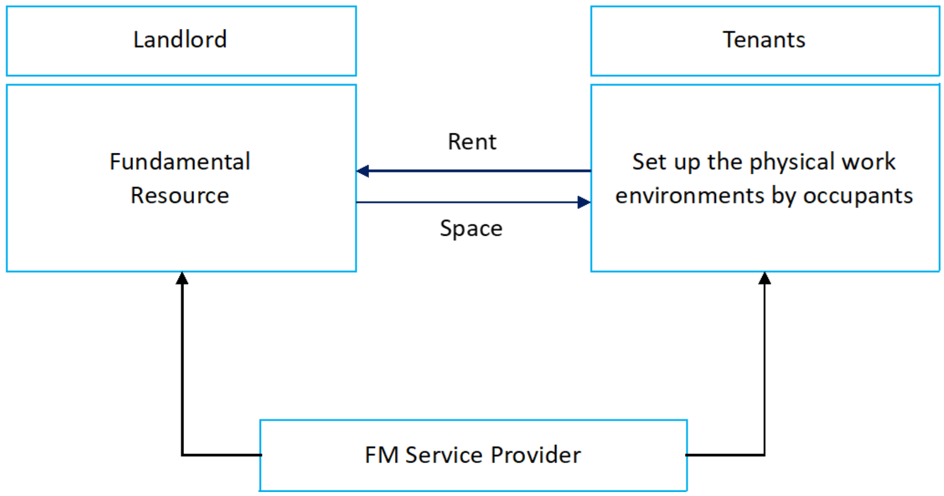

Figure 1. Conventional FM model: services supporting occupants and building owners

However, the traditional FM business model is enforced to adapt to the changing economic trends like the rise of the sharing and gig economy. The gig economy promotes independent, project-based work, and aligns small businesses with characteristics of Millennials (Petriglieri et al. 2018). Millennials define themselves as travelers who are not bound to an assigned place and do not own things except for digital nomad possessions (Goldman Sachs 2019). This culture of digital nomads promotes a sharing economy in which physical assets such as offices, houses, and cars, can be shared as services. In this context, third places such as café and libraries are preferred as workplaces by young-entrepreneurs and independent professionals (Brown 2017).

A sharing economy business model provides a hint of how the FM industry adapts to changing circumstances. There are service enablers or intermediaries between service providers and customers. Service enablers provide a platform service to connect service providers and the customers reflecting feedback from both entities for better service quality (Kumar, Lahiri, and Dogan 2018).

Since tenants are not fixed, the present views in FM need to be reconsidered depending on the changing customer needs. As the composition of building users changes, the role of FM can become a third-party space and service provider that produces a profit from anonymous building users (Figure 2). The new business model creates profit from not only providing traditional FM services such as janitorial, and operations and maintenance, but also a physical platform service providing differentiated spaces and services acting as an intermediary between a tenant and a building owner. Thus, the fitted services and physical environments can generate profits. Furthermore, the third-party service providers (including CWS providers) can influence the real estate and FM industry by being lessees and providing added value to actual space users.

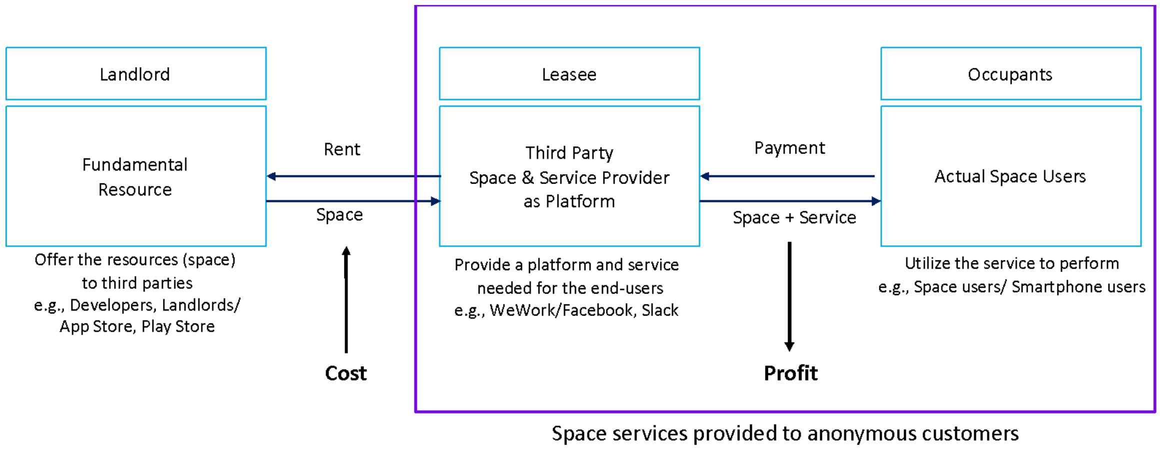

Figure 2. Redefined FM model: FM as a space & service provider

By involving the spatial and service factors in FM, the FM industry would have adaptability to the rapidly changing factors such as workforce demographics, evolving technology, and infectious disease which is then followed by changing working patterns. CWSs are considered one of the embodiments of this new conceptual model of FM. In this study, both spatial and service factors within CWSs are explored to improve the satisfaction of CWS users, which would finally lead to changes in CWSs.

CHAPTER 2. A REVIEW OF PHYSICAL ENVIRONMENT AND SERVICE IN WORKPLACE

Workplace is *'the physical settings in which work happens, to the services that support people in those settings and, perhaps most critically, the management processes that enable their effective use'* (Alexander et al. 2004). Since one's satisfaction with one's work environment is associated with job satisfaction and retention as well as individual and organizational performance (Carlopio 1996; Jain and Kaur 2014; Lee and Brand 2005), the changing demands for the workplace should be addressed in a deliberate manner.

2.1 Physical Work Environment

Indoor environment in office is associated with occupants' comfort and performance in office (Clements-Croome 2006; Al Horr et al. 2016). Indoor environment includes diverse features such as ambient environments (thermal comfort, indoor air quality, lighting, and acoustic comfort), office layouts, outside views, design aesthetics, and visual and acoustic privacy (Al Horr et al. 2016).

2.1.1 Ambient Environment

Thermal comfort and acoustic comfort more significantly influence the overall satisfaction of indoor environments when they are not acceptable whereas IAQ and lighting are linearly related to the overall satisfaction (Huang et al. 2012; Kim and de Dear 2013). Those two factors influenced the satisfaction of the other factors (Huang et al. 2012). In order to increase satisfaction, the controllability of temperature is suggested so that occupants can adjust room conditions to fit their preferences (Ng 2010).

Indoor environment satisfaction is positively associated with satisfaction with furniture design and layout that meets occupants' physical and privacy needs as well as supports tasks (Marquardt, Veitch, and Charles 2002). The adjustability and comfort of furniture and configuration are necessary to increase occupant satisfaction. Space arrangement and furniture is also considered a performance indicator in workplace (Hassanain, Alnuaimi, and Sanni-Anibire 2018).

2.1.2 Office Layout

Open-plan layouts have been adopted in many corporate offices to increase interaction (Allen and Henn 2007). Allen (1970) claimed a relationship between the number of communications and the proximity of knowledge workers as well as visibility. A close distance increased accessibility to others and open-plan layouts provided more visibility than enclosed layouts. Thus, open-plan layouts increased the opportunity for face-to-face interactions and collaboration (Hoendervanger et al. 2019). Increased interactions facilitate the exchanges of tacit knowledge that trigger innovation (Nonaka 1994).

However, there is a contrary view that the frequency of face-to-face interactions can be decreased in open-plan layouts because of occupants' concerns about acoustic privacy and subsequent reliance on e-mail or instant messages to communicate (Bernstein and Turban 2018). Another study presents that open-plan layouts entail undesirable effects such as decrease in performance, privacy, and satisfaction of indoor environments (De Been and Beijer 2014). Since visual privacy significantly influences the overall satisfaction of indoor environments, the proportion of enclosed and open-plan spaces should be deliberately planned (Kim and de Dear 2012, 2013).

2.1.3 Outside View

Views of nature and plants through windows reduce anxiety and tension of occupants in workplace (Chang and Chen 2005). The study by Chang and Chen (2005) showed that employees who can see views of nature through windows showed more psycho and physiological benefits in terms of satisfaction and stress than those who are with restricted views of nature in the workplace.

Daylight also increases the physiological and psychological benefits to occupants provided it does not cause glare (Shishegar and Boubekri 2016). Daylight contributes to providing sufficient illuminance indoors that increases a visual task performance as well as adjusting hormones such as cortisol and melatonin that are related to alertness and sleep quality and even stress level (van Bommel 2006; van Bommel and van den Beld 2004).

2.1.4 Design Aesthetics

Occupants' emotion is influenced by the color schemes and texture of the space (Kaya and Epps 2004). For instance, the color green makes the occupants feel calm and relaxed. Such cool colors also cause fewer visual distractions from the surrounding environments than warm colors such as red and orange (Kwallek et al. 1996). Office design also presents organizations' culture and values that facilitate reminding employees of the organizational values and familiarizing them with the company culture (Al Horr et al. 2016).

In addition to indoor environmental features, aesthetic design in the office promotes occupants' physical activities and contributes to occupant comfort (McArthur and Powell 2020). For instance, music and artworks provide complexity to space, which provides pleasure to occupants and encourages physical activities (McArthur and Powell 2020). Proportional ceiling height to a room size creates a different look and feel of the space such as a cozy and open feel (International WELL Building Institute 2022). Meyers-Levy and Zhu (2007) also claim that a higher ceiling is associated with a freedom-related concept (creativity), which stimulates the interdependence process of given information. This process entails uninhibited approaches to finding intersecting features among multiple data pieces.

2.1.5 Technology and Facilities

Advanced information communication and technology have enabled organizations to hire essential employees and agilely respond to rapidly changing market conditions (Townsend, DeMarie, and Hendrickson 2000). Hardware and software tools are essential to perform at work with greater quality and speed (Davenport 2005). Since the number of independent knowledge workers is increasing, wireless internet service (WiFi) is now an essential technology support and a key factor in choosing facilities (Lee 2018).

Sport facilities and recreational spaces close to office help employees easily access leisure activities and keep them healthy (Al Horr et al. 2016). The connectivity as well as the physical proximity to parks also promote physical activity (Lopez and Hynes 2006; Sallis et al. 2012). In addition to the physical environment, urban cities provide diverse recreational programs at parks and public spaces to promote increased physical activity (Floyd et al. 2008).

Childcare facilities near the office provide employees flexibility between work and childcare and help them to focus their work without the anxiety of caring for their children at the same time (Horizons 2004). The reduced anxiety and work interruptions resulting from the availability of childcare decreased absenteeism and turnover rates (Payne, Cook, and Diaz 2012). In-house childcare service increased the job satisfaction and organizational commitment (Greenberger et al. 1989). Such workplaces provide high quality and accessibility of childcare such as assisting with the search for qualified caregivers and alleviating concerns about the children's welfare (Payne et al. 2012; Skinner and Chapman 2013).

2.2 Service in Workplace

Besides physical work environment, service in workplace assists employees and workers to concentrate their core jobs as well as boost employee productivity (Dearden, Reed, and Van Reenen 2006; Lait and Wallace 2002).

2.2.1 Administrative Support

Administrative management provides service and specified support to individuals and manages information which in turn encourage employees to concentrate on their core work (Ferreira, Erasmus, and Groenewald 2010). Administrative support can reduce workers' amount of routine paperwork such as compiling information, which helps employees lower their stress from heavy workloads and concentrate on their main job (Lait and Wallace 2002).

2.2.2 Workplace Education

Krueger and Rouse (1998) investigated the effects of workplace education on earnings, turnover, and job performance of employees of a manufacturing company and a service company. Workplace education was positively associated with workers' job performance, individually as well as for groups in the service company. For instance, employees who participated in workplace education earned more awards and were nominated more than others who did not. Dearden et al. (2006) also confirmed that job training was highly associated with the productivity of employees.

2.2.3 Health Promoting Programs

Physical activity programs or interventions provided in the workplace can be effective to promote physical activity such as an email service reminding users of workout time and personal training service (Floyd et al. 2008; Malik, Blake, and Suggs 2014). Malik et al. (2014) comprehensively reviewed studies of health promotion interventions in the workplace to reveal effective ways to encourage people to perform physical activity. In the study, counselling and health promotion information and messages significantly increased the level of physical activity; for instance, tailored counselling and material interventions like mentors and customized emails increased the frequency of exercise. Team-based workout program more promoted to increase the physical activity behavior.

CHAPTER 3. PHYSICAL ENVIRONMENT AND SERVICE IN COWORKING SPACE

Coworking is a way to work by gathering individuals such as freelancers and employees of small enterprises and organizations in one place so they can share knowledge and information to solve problems and to create innovation through social interactions (Kojo and Nenonen 2014; Uda 2013). Waters-Lynch et al. (2016) positioned coworking at the intersection among creative knowledge work, independent work, and self-employment. Coworking culture is rooted in sharing economy and accelerated by the information and communication technology such as virtual meeting applications and virtual collaboration platforms (Bouncken and Reuschl 2018; Sutherland and Jarrahi 2018), which facilitate workplace flexibility sought by CWS users (Yang, Bisson, and Sanborn 2019).

Uda (2013) described coworking with two sub-concepts: coworkers and the workplace. Coworkers are autonomous individuals who are encouraged to communicate and collaborate with others. Uda (2013) illustrated the theoretical status of working individuals according to the degree of physical contact such as the frequency of communication with other people and the diversity of contact from diverse backgrounds while interacting within the workplace. Coworkers among freelancers, small-scale entrepreneurs, and organization members presented the highest level of diversity while working with others and a high level of contact with others.

A CWS is a workplace where autonomous individuals or a group of individuals pay monthly fees to utilize not only the space and its facilities, but also socialization services

(Bouncken and Reuschl 2018; Uda 2013; Waters-Lynch and Potts 2017). CWSs provide opportunities to network with others and form communities based on 'task-related targets' and 'leisure targets' depending on the autonomy of the space users (Bouncken and Reuschl 2018). The role of CWSs is to promote networking for knowledge transfer between various knowledge workers and resolve problems in businesses.

There are similarities between service facilities and CWSs. Service facilities are considered an older version of CWSs that provide facilities, amenities, and services to support the space users' professional activities. However, while service facilities are satellite or dispersed offices from central offices for a strategic purpose (Kojo and Nenonen 2017), CWSs bring informal, communicative, and creative culture into an organizational culture and the workplace (Waters-Lynch et al. 2016).

COVID-19 forced conventional corporate employees to work from home and the prolonged situation shifted office workers' perception of working full-time in an office. The majority of employees do not want to go back to work in one office full-time (Slack 2020). Workers want work environments similar to where they are used to working (Ng 2010). However, home offices are not enough to fulfill the functional and cultural needs of most workers. For example, in a traditional office, work-related facilities and services such as amenities and administrative services as well as networking opportunities have been provided by organizations. Information coordination (Waters-Lynch and Potts 2017) and social isolation (Raffaele and Connell 2016) issues of the self-employed and employees who work in dispersed workspaces should be resolved.

From the aspect of facility management, facility management companies only needed to operate in physical environments to support the core business of tenants pursuant to the contracts with the building owners or the organizations (Atkin and Brooks 2015). However, with the changing needs of workplace flexibility (Smith et al. 2020), the role of facility management is expanding to workplace management that provides and manages not only the physical workspace, but also provides services that support work for professional individuals as a third place (Brown 2017).

As the demand for workplace changes due to changing environmental factors, the CWS is emerging as an attractive alternative workplace. After the COVID-19 outbreak, work flexibility, and suburbanization, CWSs are attracting startup and small-to-medium (SME) employees as well as traditional office workers who prefer to work closer to home for better performance (Smith et al. 2020). In this context, CWSs are a form of alternative offices generated by knowledge workers' preferences according to changing work patterns (Waters-Lynch and Potts 2017).

3.1 Physical Work Environment in Coworking Space

3.1.1 Ambient Environment

CWS users choose CWSs for networking and a creative atmosphere that encourages frequent communications (Bouncken and Reuschl 2018; Weijs-Perrée et al. 2020). Thus, such workers are expected to present different responses to ambient environments, for instance, noise from communication. Lee (2018) conducted research into the motivations and preferences for environmental features of CWSs. The survey results presented that physical workspace factors such as the environmental quality of the space was the second motivator for choosing CWSs out of four primary factors. The gap analysis in the study showed that the control of acoustic privacy and thermal comfort were the features that most often did not meet expectations. Orel and Almeida (2019) highlighted that closed meeting rooms are necessary to secure acoustic and visual privacy.

3.1.2 Space Layout

CWSs are open to visitors and space users to stimulate informal and formal communication (Yang et al. 2019). Such open concept design with openness and proximity promotes creativity through active interactions. When employees were close to each other, they had more communication with more chances to bump into each other (Allen 1970, 1977). Sanborn (2015) indicated learning advantages of overhearing other coworkers' conversations in close proximity. Knowledge workers also had more interactions when colleagues were within their sights or were passed frequently, for example, off common corridors (Rashid et al. 2006).

The space setting within CWSs are mainly divided into two settings: space for work and space for amenities (Han 2013). CWSs provide diverse space configurations including activity-based working space, open-plan offices, concentrating rooms, phone booths, meeting rooms, tech sectors, amenity areas, lounges, and kitchens (Sanborn 2015; Yang et al. 2019). Some CWS organizations provide a break room as well as a personal training space (e.g., WeWork).

3.1.3 *Interior and Aesthetics*

The design aesthetics of CWSs was highly ranked as the third most important attribute out of fourteen (Lee 2018). Customized spaces generate an atmosphere that promotes socialization and collaboration with an informal and communicative culture (Waters-Lynch et al. 2016). The aesthetics of space present the embedded culture of space users and user groups (Liegl 2014). The design of CWSs should stimulate the creativity of users. Color schemes and the texture of the walls and furniture create different moods for a space. For instance, various colors are used in CWSs to make such spaces vibrant as well as meet the different preferences of users (Orel and Almeida 2019).

3.1.4 *Technology and Facilities*

The study by Lee (2018) indicated that the highest-ranked motivator to use CWSs was technology support such as wireless internet service. Since self-employed individuals or entrepreneurs, and freelancers from creative industries are the majority user groups accounting for 70% of CWS users (Foertsch 2011; Spinuzzi 2012), technology support including WiFi connectivity are essential for such alternative arrangement work populations who work through online platforms and virtual meeting applications (Mas and Pallais 2020). Other tech amenities are offered in CWSs such as audio and video conferencing equipment and display monitors.

Besides technology infrastructure, CWSs have not only eating and break-out spaces, but also recreational and leisure purpose facilities including on-site gyms, ping-pong tables, yoga rooms, nap zones and spas (Waters-Lynch and Potts 2017; Weijs-Perrée et al. 2020). Although these facilities are considered to increase knowledge workers' performance, the facilities should be configured in a way so as not to distract other space users (Button 2019).

Some CWSs provide childcare facilities (Johnson 2018; Warkentin 2020). Orel (2019) introduced CWSs as an optimal workplace that resolves work and family conflicts. The author observed that some CWSs supported space users’ tasks by allowing them to bring their kids to the workspace and to combine their obligations at work and childcare. van Blokland (2018) highlighted the importance of childcare services in a CWS to attract customers. For instance, a CWS can support working parents by not only providing professional childcare service (Johnson 2018), but also generating communities in the CWS to care for their children (Orel 2019).

3.2 Service in Coworking Space

3.2.1 Administrative Support

A CWS is to provide administrative services. It is similar to the subscription model of a platform service. In the context of CWSs, a platform service provider manages office chores and maintenance of facilities. In addition, a third party that has a CWS membership or a partnership with a CWS company provides administrative services for freelancers and SMEs (Cabral and Winden 2016).

3.2.2 Workplace Education

Some CWS companies have a partnership with professional job education companies. Cabral and Winden (2016) mentioned that CWSs can provide educational programs, and workshops to promote interactions between coworkers. Also, freelancers and employees of SMEs can improve their job performance with educational programs (Krueger and Rouse 1998).

3.2.3 Socialization and Networking Service

One major difference of CWSs from traditional offices is the existence of community managers or CWS hosts (Brown 2017; Merkel 2015). They operate the facilities as well as host inspirational and networking events to encourage space users to have more social interactions (Merkel 2015). As Bernstein and Turban (2018) insist, active collaboration and interactions are not simply depending on proximity and openness, enrollment is a crucial factor for collaboration. Thus, the role of coworking hosts as communication facilitators is magnified (Merkel 2015). Brown (2017) evaluated the effects of community managers as a mediator of relationships between coworkers and interactions. The study presented that, except for the 'reluctant soloist', the efforts of community managers to connect space users and encourage interactions and collaboration were effective and even increase mutual trust among coworkers. Thus, the degree of collaboration can be improved by 'visionary' managers (Liimatainen 2015; Merkel 2015). Moreover, the increase in interactions between CWS users encourages the creation of a community that alleviates the isolation of independent workers (Surman 2013).

Interactions can be also facilitated by the operation policy of a CWS. CWS firms have different policies to rent or provide the space, which can create diversity in the types of CWSs. Some of them provides 'drop-in' service that if people who do not regularly use a designated CWS want to use a certain CWS, they can use the space with flexible payment options such as one day or hourly payments (Garrett, Spreitzer, and Bacevice 2017; Uda

2013). If CWS tenants have a meeting with their partners, the partners can visit the CWS for the meeting. Not only internal but also external events can taken place. This draws external professionals to the space. This context presents that the operation policy can adjust the openness and how much CWS users mingle with others. Since infusion of knowledge from a variety of sources is important to create innovation (von Hippel, 2001), Liimatainen (2015) highlights the diversity in the composition of actors. The consideration of the proportion of actors from different industries and professionals may help space users increase their knowledge productivity.

3.2.4 Leisure and Well-being

CWSs provide activity programs not only including networking events, but also other physical and leisure activities (i.e., wine, coffee, yoga, meditation classes, etc.) (Weijs-Perrée et al. 2020). Several studies consider these activities community-building activities that facilitate creativity of members and which form a creative culture (Cabral and Winden 2016; Mariotti, Pacchi, and Di Vita 2017). The community-building activity fosters interactions which lead to reduced social isolation and managing mental health (Mariotti et al. 2017; Merkel 2015). In fact, workers who work from home or from coffee shops presented negative feedback such as distractions, self-motivation problems, and a sense of isolation in interviews (Spinuzzi 2012). In addition to mental health, programs promoting physical activity such as ping-pong, yoga, Pilates, and cardio programs can be designed for CWS users according to space capabilities (Gunawan 2018).

3.3 Challenges in Coworking Space Research

Although CWSs consist of providing diverse facilities and services, the current literature lacks studies on user preferences for the facilities and services (Appel-Meulenbroek et al. 2021; Weijs-Perrée et al. 2020). The previous studies of CWSs have focused on collaboration and social capital, and spatial factors in relation to overall satisfaction and collaboration (Lee 2018; Sanborn 2015; Yang et al. 2019). Regardless of the motivations of CWSs, commercial (for-profit) to community-based non-profit entities, understanding user preference is critical for the effective operation and sustainability.

Understanding user preference became more important due to the impact of COVID-19 since the way of working and collaborative as well as expected social etiquette have dramatically changed over the short period of time. This pandemic has also increased private office and sanitary service needs (Smith et al. 2020). The changing public behaviors and demands require to investigate changed preferences for facility and service factors within CWSs.

Besides the gaps identified in the literature review, most of the previous physical workplace management research often based on corporate workplace where users may not have a direct power on selecting their physical workplaces. Whereas, CWS users are the individual members or small start-up companies who pay for CWS membership for a short-term basis as short as a day. This volatility of selecting CWSs within the same region and the business model of CWSs give a lot more decision-making power for space users compared to users in corporate workplaces. These direct payment and short-term commitment create users' full perception of them being a client of CWSs and they become more enthusiastic for sharing their opinions and giving feedback through online platforms, such as social media. Such social media data as e-word of mouth are invaluable in analyzing the satisfaction with facilities and services of CWSs.

This study employs web scrapping to collect data of user satisfaction within physical workplace leveraging the attributes of CWS, whereas the previous studies have investigated user satisfaction within physical workplace through surveys, interviews, and observations. In order to analyze the big and unsolicited text data, analyzing metadata such as categorical information is essential to understand the voluminous data and achieve a data-driven decision making (Michener 2006). Categorizations are suggested to analyze the unsolicited and user-generated social media data effectively and efficiently, which leads to concluding general preferences for CWS factors. With grasping the general view, maintenance data and customer service records in a CWS company can reveal the performance of a specific CWS in detail. Since such records consist of data in text and unorganized formats, computational processing that presents meaningful information is needed. In this context, Natural Language Processing (NLP) and machine learning are suggested.

NLP transforms text information into numeric information based on linguistic and technical rules to enable computers to understand text information (Liddy 2001). In particular, machine learning based approaches made a big step to achieve ‘human-like language processing’ by a computer (Devlin et al. 2018). NLP facilitates a computational analysis of massive user-generated text data stored in such online platforms. Feature engineering and extractions in NLP present important words, phrases, and sentences while excluding meaningless words, which helps build a high accuracy machine learning model (Zhai et al. 2018).

In addition to the improvements of the NLP performance by machine learning itself, machine learning performs diverse functions to analyze data and generate information through classification, clustering, and prediction (Rebala, Ravi, and Churiwala 2019). Multiple algorithms are utilized to deliver those functions and examined to find the best performing algorithms because each model has different bases from statistics to artificial neural network (Bishop 2006). Thus, appropriate machine learning techniques should be explored to analyze the big data collected from an online platform and examined the performance.

3.4 Overarching Research Objectives

The research gaps discussed in the review give rise to the following research objectives:

1) identify preferences for facilities and services factors in CWSs during COVID-19; 2) identify changing preferences for factors about facilities and services during COVID-19; 3) propose the applications of machine learning and NLP techniques and demonstrate the applicability of computational data collection and analysis methods in the physical workplace management research.

Based on the research objectives, three overarching research questions are formulated. 1) What are critical factors in facilities and services associated with user preferences in CWSs during COVID-19? 2) Which demands specific facilities and services are changed comparing between prior to COVID-19 and during COVID-19 periods in

CWSs? 3) How to apply NLP and machine learning techniques in analyzing social media data in relation to user experience in CWSs as well as data about maintenance and customer complaints?

This research consists of two studies shown in Figure 3. Study I delivers thematic categories of facility and service factors. The thematic categories are offered by the literature review of CWS. In order to provide general user preferences for the facility and service factors, the social media data in relation to CWS in multiple cities are collected and analyzed based on the categories to achieve the research objectives 1 and 2 utilizing NLP and machine learning techniques.

Study II proposes the integration of data from different sources (software) to enable a single or small CWSs to build a robust machine learning models, to automate a data management process, and to finally gain meta information according to specific data management needs. Study II examines facility maintenance, for instance, Mechanical (HVAC), Electrical, and Plumbing (MEP) issues. MEP issues are directly related to CWS users' satisfaction such as thermal and lighting comforts, and the convenience of using power outlets, rest rooms, and kitchenette. Thus, the maintenance issues can be used to measure the performance of CWS in the aspect of hard issues. The facility maintenance requests in CWS are managed by a FM department in the building. Study II utilizes a maintenance data of a building that includes a CWS.

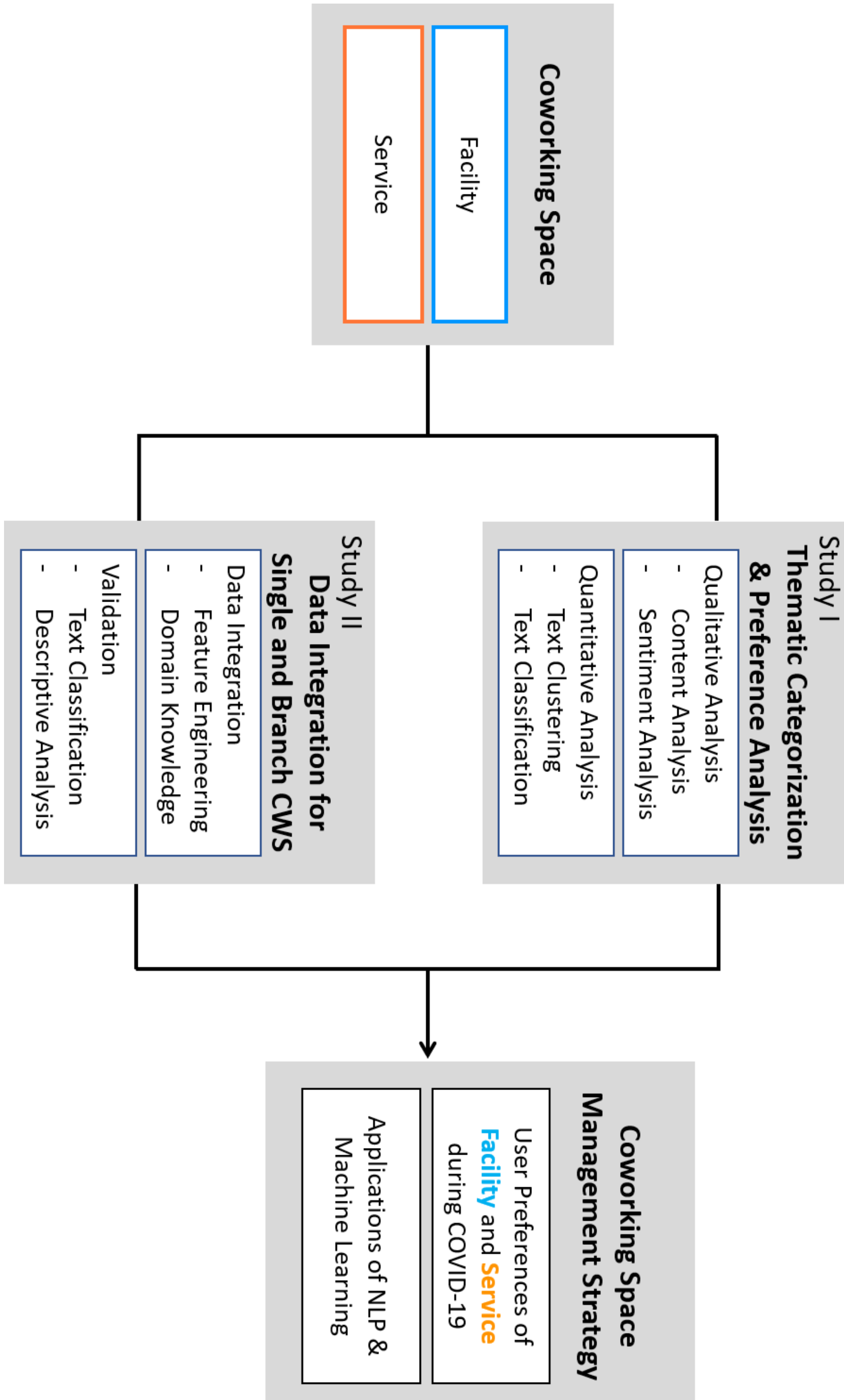

Figure 3. Overview of studies to build CWS management strategies

CHAPTER 4. STUDY I: THEMATIC CATEGORIZATION AND USER PREFERENCE ANALYSIS OF COWORKING SPACES DURING COVID-19 USING SOCIAL MEDIA DATA

4.1 Introduction

A “third place” has become popular with the emergence of the gig economy and new working patterns such as nomad-style work and self-employment. A third place is a public place outside "the first place, the home” and “the second place, the workplace,” which promotes interactions and socialization in the community (e.g., coffee shops, libraries, and community centers) (Oldenburg and Brissett 1982). A CWS is one of the types of a third place; it offers physical environments and various services to accommodate the users’ work. The services provided by the CWS function possibly influence the user experience as the physical environmental satisfaction is intertwined with management supports (Moezzi and Goins 2011). Previous studies have investigated preferences for and satisfaction with CWS factors to reveal motivators that encourage people to use CWS, such as staff (community managers), space, technology, service, accessibility, and membership (Appel-Meulenbroek et al. 2021; Lee 2018; Weijs-Perrée et al. 2020).

Community managers (staff) in CWS provide hospitality services to assist in using the physical environments and the services. Especially, their commitment to facilitating interactions among CWS users is essential. The community managers host networking events and coordinate the interactions between coworkers who are not necessarily working together because serendipitous encounters hardly occur without intervention or coordination by CWS managers (Brown 2017; Liimatainen 2015). The community managers' curation is necessitated for meaningful coworking such as knowledge sharing and collaboration.

Spatial factors consist of multiple components such as ambiance, aesthetics with the atmosphere (vibes), and the status of facilities, which are related to the users' experience in CWS. For instance, ambient environments such as thermal, acoustic, visual comfort, and privacy are positively associated with satisfaction in CWS (Lee 2018). Space configurations include working space and service sectors like meeting rooms, kitchens, and amenity areas (Sanborn 2015; Yang et al. 2019). Interior and aesthetics provide a professional and socialization-promoting atmosphere that is one of the purposes of using CWS (Endrissat and Leclercq-Vandelannoitte 2021; Waters-Lynch et al. 2016; Weijs-Perrée et al. 2019). The status of facilities, such as the cleanliness of spaces and restrooms and availability of meeting rooms, are factors in choosing CWS (Waters-Lynch and Potts 2017; Weijs-Perrée et al. 2020).

Technology is a critical factor in CWS, which enables coworkers to be free from spatial and temporal constraints to work and conventional work arrangements (Endrissat and Leclercq-Vandelannoitte 2021). Therefore, stable internet connections (i.e., wired and wireless internet) and equipment for video conferencing and virtual meetings should be provided to support flexible and remote formats in work (Lee 2018; Mas and Pallais 2020).

Services in CWS support various necessary activities occurring in CWS. First, CWS provides work support-related services such as childcare services (Johnson 2018; Orel 2019), administrative-like virtual offices, and professional job education services provided by a third party (Cabral and Winden 2016). In addition, community socialization activities and networking events are organized by CWS managers to promote interactions among coworkers as well as outside members (Spreitzer et al. 2017; Uda 2013). Refreshments such as coffee and snacks are included accordingly in service sectors (Morisson 2018).

Diversity of members in terms of professional positions is managed for productive interactions for their businesses (Liimatainen 2015). Extra services unrelated to work are programmed for CWS members, such as leisure and well-being activities that address physical and mental health, for example, ping-pong, yoga, and meditation classes, to wine and coffee classes (Waters-Lynch and Potts 2017).

General real estate factors such as accessibility, location, and lease or membership contracts are also important motivators for choosing CWSs (Appel-Meulenbroek et al. 2021). Users prefer CWSs close to public transportation and which are located downtown. Convenient locations provide easy access to the members and the clients (Bouncken and Reuschl 2018; Waters-Lynch and Potts 2017). According to CWS policies, varied memberships exist from one-day passes to long-term contracts (Uda 2013). The various membership contracts offer additional options to accommodate CWS users' needs.

CWS users may have different individual purposes in using a space as their workplace. According to the specific purpose, their needs may vary, such as the need for a private office, communal space, technology, and networking opportunities with others (Kojo and Nenonen 2017; Lee 2018; Spinuzzi 2012; Yang et al. 2019). In addition, responsive and adaptive behavior to COVID-19 aligns with the needs of CWS managers and users. For example, people are reluctant to sit, and constrained from sitting, close to one another and to work with others due to infection risk, although physical proximity is a key concept of CWSs. Therefore, the demands for spaces for meetings and events considerably decreased (Mayerhoffer 2021). On the other hand, distance restrictions between the space users and high hygienic standards may lead to changes in the operational strategies of CWSs. In this context, questions arise on the user experience of CWS factors in the U.S. and the change in users' preferences for CWSs due to COVID-19.

Interview and survey formats have been mainly utilized to examine occupants' comfort, satisfaction, and interactions within workplaces to address research questions about physical workplaces. Data sharing of these types of data is not preferred by and considered extra work for researchers due to the difficulty in collecting and integrating the data (Kleppner 2010; Tenopir et al. 2011). Thus, the limited responses lead to difficulty in demonstrating general indoor environmental effects. Indoor environmental quality (IEQ) survey data systematically accumulated for various buildings by the Center for the Built Environment (CBE) is only a massive database stored and utilized to gain a general idea of IEQ in physical workplaces (Kim and de Dear 2012; Moezzi and Goins 2011).

Thanks to diverging types of workplace, including multi-tenant workspaces or flexible offices (e.g., CWSs) (Kojo and Nenonen 2015, 2017), open user-created data about physical workplaces are available through multiple digital platforms, such as Yelp reviews, Twitter, and Instagram. Anonymous space users and contracted personnel share their experiences of spaces and services in CWSs via an online platform. Massive amounts of data on such workspaces are published through online platforms called e-word of mouth (EWOM), such as Yelp, Twitter, Square, and Swarm. The provided services and the information on physical environments are considered important to appeal to customers (i.e., knowledge workers). The information exchange among customers via an online platform incrementally influences consumers' decisions (Jones, 2009). A massive volume of sentiment data is created by users, such as opinions and emotions of products and services (Kharde & Sonawane, 2016a). Thus, the analysis of EWOM could provide a general guide for CWS design and operation planning.

Moezzi & Goins (2011) applied text-mining techniques in the open question responses accumulated by CBE to draw general complaints about IEQ in physical workplaces. Villeneuve & O'Brien (2020) extracted IEQ data of residential buildings from AirBnB data sets to explore IEQ satisfaction. The results showed that user complaints made up 7% of the total comments and 70% of these complaints were IEQ complaints. Among the IEQ complaints, 60% was acoustic-related, 26% was thermal-related, 10% was IAQ-related, and 4% was visual complaints. Chinazzo (2021) investigated job reviews published on Glassdoor to extract IEQ-related information in workplaces. The extracted data in the study (Chinazzo 2021) were analyzed through rule-based approaches and an iterative cleaning process. Chinazzo (2021) and Villeneuve & O'Brien (2020) generated customized IEQ-related word sets and selected comments, including such word(s) in the customized sets. Words around IEQ terms were also extracted to understand the context of the sentence in the studies. To gain better results, iterative data cleaning and selection of relevant and irrelevant words were performed.

Rule-based classification requires manual and repetitive word selecting procedures until satisfactory results are achieved. Therefore, the rule-based classification process is time-consuming and includes human error and less scalability than a machine learning classification process (Cronin et al. 2017). Using deep learning methods, online data can also be analyzed through a clustering process to extract important attributes embedded in open text responses. Due to the rapid improvement of Natural Language Processing (NLP), techniques using deep learning, classification accuracy, and clustering quality have strikingly increased considering the sentence context (Devlin et al. 2018).

This study aims to 1) explore the user experience of CWS factors in the U.S., 2) identify users' preferences for CWSs during COVID-19, and 3) demonstrate the applicability of new mixed methods in this type of research. The adopted mixed-methods include content analysis, machine learning, and transformer-based deep learning, which allow classification of open-text responses to CWS and extraction of critical features pertaining to user satisfaction in general. Model validation is conducted with the results. This study addresses the following four research questions:

RQ1. What are the reasons for (dis)satisfaction with the physical environment and service factors in CWS during COVID-19?

RQ2. Are there changes in user experiences with CWS from before and during the COVID-19 pandemic?

RQ3. Does deep learning-based transformer clustering provide theoretical insight into the characterization of a CWS users' experience?

RQ4. Do deep learning transformers and machine learning models perform with high accuracy to justify automating the classification of CWS experience reviews?

4.2 Methods

The methods consisted of three phases: 1) data collection and cleaning, 2) text data clustering and classification, and 3) validation and analysis (Figure 4). First, data collection and cleaning were conducted through web scraping. Second, the collected data was processed through three pillar methods: 1) content analysis, 2) supervised (classification), and 3) unsupervised (clustering) machine learning methods. The mixed-method approach aimed to reveal the data set's overview, identify each review's category or categories, and critical features of CWS user experience within the reviews. An issue with unsolicited and open text responses was having multiple topics in one review; for instance, one review included feedback on staff's hospitality, IEQ, and membership contracts. Sentence-level clustering and multi-label classification methods were performed to address this issue. Third, the machine learning model was validated, and the results were interpreted to prove the concepts of the machine learning and deep learning applications compared with the content analysis results.

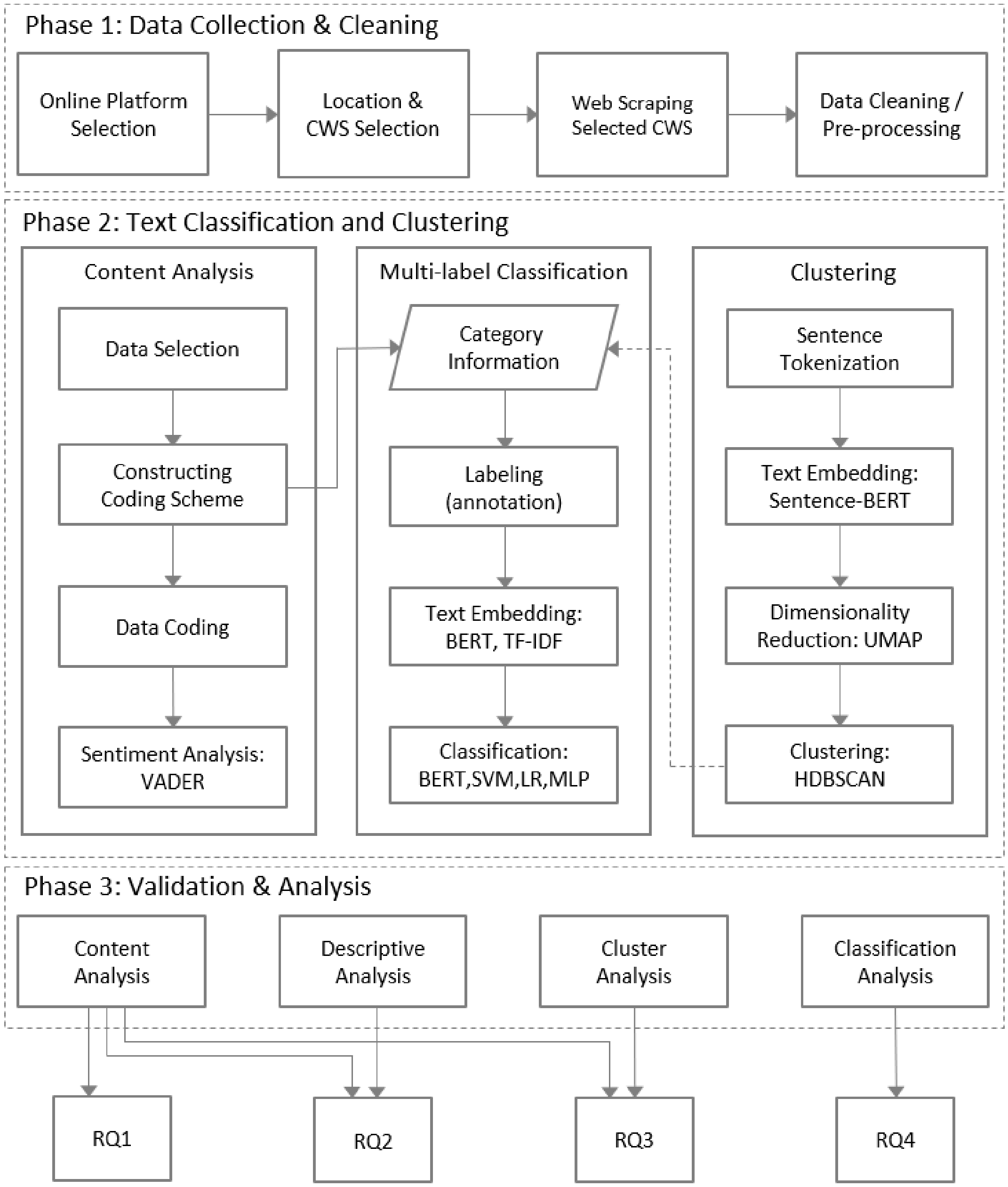

Figure 4. The flow of text mining and analysis

Based on the literature review, thematic categories of CWS were introduced for the content analysis and classification. There were six categories: staff (community managers), space, technology, service, accessibility, and membership, with twenty sub-categories

(Table 1). COVID-19 and its impacts on the six thematic categories were not included as a category because the collected reviews before the pandemic did not include COVID-19-related content. The physical work environments were broken down into three factors: space, technology, and accessibility. The categorizing schemes are described with examples in Appendix A.1.

Table 1. Thematic categories of CWS

Categories (6)	Sub-categories (21)
Staff (community managers)	Hospitality, work support, community support, maintenance of spaces (e.g., cleaning, organizing, and repairing) (4)
Space	IEQ, space configuration, interior design, atmosphere, facilities (5)
Technology	Internet access, printing service, virtual conferencing system, booking system (4)
Service	Networking and community events, work supportive service (childcare and administrative service), leisure programs, refreshment (4)
Accessibility	Location, parking lots (2)
Membership	Membership types, contract period (2)

4.2.1 *Data Collection and Cleaning*

Figure 5. Locations of the Yelp reviews collected

In Phase 1 of this study, the researchers selected an online platform, *Yelp,* and searched for CWSs on the platform. First, the researchers reviewed the searched CWSs in the platform to screen invalid results. Next, web scraping was performed of the valid CWSs, such as the name of CWSs, users, dates, ratings, and reviews. The BeautifulSoup library was utilized for web scraping in Python. A total of 4,254 reviews of 543 CWS were collected on July 23rd, 2021, centered on 11 cities in the U.S., including Atlanta, GA, Austin, TX, Chicago, IL, Denver, CO, Houston, TX, New York City, NY, Los Angeles, CA, Philadelphia, PA, Phoenix, AZ, San Diego, CA, San Francisco, CA. The period of the collected data, Yelp reviews of CWSs, was from September 2006 to July 2021. The search term was 'CWS.' In the Yelp category, 'Shared space' fitted the searching intention. Irrelevant results to CWSs and results without a review were screened out. One invalid data point, composed of only unreadable symbols, and nine duplicated data points were removed during the data cleaning process, resulting in 4,244 for the next step. Words or phrases consisting of meaningless symbols and digits were cleaned. After the data collection and cleaning steps, the collected reviews were pre-processed through lowering and lemmatizing words and removing stop-words. Lemmatization converted the words into an original form, for instance, 'studies' converted into 'study.' Stop-words included articles, conjunctions, prepositions, and pronouns.

4.2.2 Mixed-methods Approach: Content Analysis, Classification, and Clustering

4.2.2.1 Content Analysis with Sentiment Analysis

Content analysis was conducted among the data set during COVID-19; 455 reviews of 217 CWSs were written after February 2020. NVivo12 was used to code and analyze the content. Phrases and sentences were coded according to relevant topics based on the categories (Table 1. Thematic categories of CWS) and COVID-19. The details of the category coding schemes were documented, including categories, definitions, example words and phrases, and example sentences in Appendix A.1. For instance, the review sentence, "Staff were extremely friendly and accommodating, and were super helpful in explaining their COVID protocols," was coded as 'Staff' and 'COVID-19.'

The coded expressions were analyzed using Sentiment Analysis. Sentiment Analysis is devised to triangulate qualitative and quantitative analysis of customer satisfaction using machine learning or lexicon and rule-based algorithms (Rambocas and Gama 2013). Sentiment analysis is a process to automatically analyze the level of customer satisfaction based on online reviews and opinions created by consumers using NLP (Batrinca and Treleaven 2014; Kharde and Sonawane 2016). The evaluation of emotion presents as polarity of emotions such as positive or negative. Many fields have utilized sentiment analysis techniques to measure customer satisfaction with products and services such as tourism and hospitality (Alaei, Becken, and Stantic 2019; Duan et al. 2016; Philander and Zhong 2016), online markets (Bhatt et al. 2015), and education service (Altrabsheh, Gaber, and Cocea 2013; Lamba and Madhusudhan 2018).

Valence Aware Dictionary for Sentiment Reasoning (VADER) in the Natural Language Toolkit (NLTK) in this study was used in this study presenting sentiment polarity from -1 to 1. Zero refers to neutral sentiment, values greater than 0 mean positive emotion, and values smaller than 0 mean negative emotion. The example is shown in Figure 6. VADER presented a high accuracy (F1=0.96) of sentiment analysis and was tuned for short social media postings on social media data (i.e., Tweets) (Gilbert and Hutto 2014).

Document:
'WFH isn't the future. Remote is the future and the ability to work from anywhere and anytime. Full flexibility, including working from offices/coworking spaces.
I've worked 10+ years remotely, and most of them outside of my home, because I like to separate work from home life.'

Sentiment Analysis by VADER:
{'neg': 0.0, 'neu': 0.851, 'pos': 0.149, 'compound': 0.7351}

Figure 6. Sentiment analysis example using VADER

4.2.2.2 Descriptive Analysis

In addition to the content analysis of the data during COVID-19, descriptive analysis of both data before and during COVID-19 was performed through labeling the category(-ies) of each review to identify changing user experience in CWS. 1,216 Yelp reviews were manually labeled in total, where 761 out of 1,216 reviews were randomly selected before the COVID-19 period (February 2020), and 455 reviews were the total reviews of CWSs during COVID-19. The category labeling schemes (Appendix A.1) were used to label reviews such as staff (CWS managers), space, technology, technology, service, accessibility, and membership (Figure 7).

Review

I really enjoyed using this space for production. It seems fairly new and is very specious. We got an official tour and they have so much to offer Atlanta entrepreneurs. The suites have all the amenities that business owner could ever want or need. If you are seeking workspace near the city, this is perfect spot.

Staff	Space	Technology	Service	Accessibility	Membership
0	1	0	0	1	0

Figure 7. Multi-label example of a CWS Yelp review

The categorized reviews were analyzed based on the dates and categories. The quantitative analysis showed changing experiences and responses in using CWSs during COVID-19. First, the proportion of each topic presented the changes in the reviews between the different periods. Second, co-occurring categories in a review were presented from the 1st rank to the 3rd rank in the different periods. This showed which category was mentioned with another topic in a review together.

4.2.2.3 Machine Learning: Clustering

Each cleaned review was tokenized into a sentence level for clustering. Since each Yelp review was likely to include multiple aspects in one review, input sentences were tokenized for accurate clustering—for example, the comment, 'Great work environment. 'Industrious' is comfortable, friendly, and modern. We are happy to have found the property when we did, realizing it was the perfect solution to accommodate our startup needs, with plenty of capacity to handle us as we grow. The location is centralized and ideal, and for those in the business of client work, you will love having in-person meetings with some stellar views of the city within the hottest parcel in Atlanta - Ponce City Market,' included terms related to 'Space' and 'Accessibility.' This long comment was tokenized into four sentences considering the punctuation. Through the tokenization process, 4,244 reviews were broken up into 32,783 sentences. First, the tokenized sentences were pre-processed. Then, the pre-processed sentences were vectorized (embedded) for the following clustering process using SentenceTransformers (SBERT) (Reimers and Gurevych 2019).

The dimensionality of the vectorized sentences was reduced using the Uniform Manifold Approximation and Projection for dimension reduction (UMAP) (McInnes, Healy, and Melville 2018) before feeding the data into a clustering algorithm. Dimensionality reduction is essential to resolve high dimensional computation burdens and extract critical attributes instead of removing undesired attributes (Van Der Maaten, Postma, and Van den Herik 2009). Cosine distances from different embedded words or sentences are suitable for text similarity (Han, Kamber, and Pei 2012). Since UMAP uses a cosine distance as a criterion to reduce the dimensions as well as present the global structure of data (McInnes et al. 2018), UMAP was used in dimensionality reduction. After that, the dimension-reduced data points were clustered by the Hierarchical Density-Based Spatial Clustering of Applications (HDBSCAN) (Campello, Moulavi, and Sander 2013; McInnes, Healy, and Astels 2017). HDBSCAN is an improved model of DBSCAN. DBSCAN is a non-parametric clustering method and clusters data points according to their density. By building a hierarchy, clusters that did not meet the required minimum number of data points were pruned so that HDBSCAN could present a clearer view of the clustered data by differentiating noise in the data set. The minimum number was heuristically determined to be 15 in tuning the parameter and reviewing the clustering results.

The tokenized sentences were grouped according to their attributes through the clustering procedures. First, the features of each group were extracted using Term Frequency-Inverse Document Frequency (TF-IDF) to identify the group attributes with bi-grams (two sequential words) (Ramos 2003). Next, the attributes were reviewed for whether the thematic categories derived from the literature reviews did not holistically cover the CWS experience. When a new thematic category was discovered, it was added to the existing categories.

4.2.2.4 Machine Learning: Multi-label Text Classification

Since a new category was not discovered in the clustering process, the labeled 1,216 reviews in the descriptive analysis step were adopted to train machine learning models. Four classification models were utilized in this study: Bidirectional Encoder Representations from Transformers (BERT), Support Vector Machine (SVM), logistic regression (LR), and Multi-layer Perceptron (MLP). The vectorization of the reviews was performed by TF-IDF and BERT embeddings. Through TF-IDF, the vectorized reviews were plugged into SVM, LR, and MLP. The deep learning-based transformer model, BERT, was utilized for both vectorization and classification. The multi-label classification problem was transformed to binary relevance using SVM, logistic regression, and MLP. Finally, sigmoid activation was applied in the outputs from BERT to calculate the probabilities. For example, suppose the probability of ‘Space’ was greater than 0.5. In that case, the sigmoid function assigned the review as ‘Space.’ This assignment was conducted on all categories because of the multi-class characteristics of the data. Hyperparameters and optimizers for each algorithm were provided in Appendix A.2.

4.2.3 Model Validation

The performances of SVM and logistic regression, MLP, and the state-of-the-art model, BERT, were evaluated and compared. The labeled data set was divided into training and test sets in a ratio of 0.75 to 0.25. The test data set was not used in the training steps. Instead, randomly selected train and test data sets were evaluated to mean accuracy values and micro-averaged f1 scores. Applying the machine learning models in the test set was evaluated through standard metrics for multi-label classifiers such as micro averaging f1 score, Hamming loss, and the overall accuracy as seen in Equation (1) to (4) for the multi-label classification. f1 score balances precision that was a True Positive (TP) over total predicted positive and recall that was true positive over total actual positive for each label. However, micro averaging f1 score calculates all the TP and all errors such as total False Positives (FPs) and total False Negatives (FNs) rather than looking at each label. TN stands for True Negative.

Equation 1. Accuracy

$$\text{Accuracy} = \frac{TP + TN}{(TP + FP + TN + FN)}$$

Equation 2. Micro F1 Score

$$\text{Micro F1 Score} = 2 \cdot \frac{Micro\text{–}precision \cdot Micro\text{–}recall}{Micro\text{–}precision + Micro\text{–}recall}$$

$$where\ Precision = \frac{TP}{TP + FP}\ and\ Recall = \frac{TP}{TP + FN}$$

Equation 3. Hamming Loss

$$Hamming\ Loss = \frac{1}{|N| \cdot |L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} xor\left(y_{i,j}, z_{i,j}\right) \ (3)$$

where N is the size of sample and L is categories,

Equation 4. Overall Accuracy

$$\text{Overall Accuracy} = 1 - \text{Hamming Loss (4)}$$

4.3 Results

4.3.1 Content Analysis with Sentiment Analysis

The results of the content analysis are summarized in Tables 2 and 3. Two aspects of user experience in CWS have been considered, overall user experience and COVID-19-relevant user experience (i.e., reviews mentioning ‘COVID-19’, ‘pandemic’ or ‘infectious disease’) of each category during the COVID-19 period. General user experience in CWS during COVID-19 is presented in Table 2. The number of coded sentences and phrases in the reviews are in order of staff (336), service (270), space (270), membership (139), accessibility (81), and technology (67). About 90% of service, space, and accessibility reviews were positive, while staff and membership showed 80% and 54% positive reviews, respectively. The highest number of negative reviews on membership was due to membership cancellations or pauses during the lockdown period. Additional charges for using facilities and services were also another reason for low satisfaction rates. In the negative experiences with staff, delayed communication or non-responsiveness to a request was rated low.

The sentiment analysis quantifies the content analysis results with positive, neutral, and negative values (i.e., 1 to -1) (Table 2). The ratios of each category are in order of positive, neutral, and negative sentiment polarity: 1) staff (79%, 4%, 17%), 2) space (94%, 2%, 4%), 3) service (83%, 13%, 4%), 4) technology (66%, 31%, 3%), 5) accessibility (75%, 19%, 6%), and 6) membership (52%, 19%, 29%). The averaged ratios of all categories were 75% (positive), 15% (neutral), and 11% (negative), respectively.

Table 2. Overall user experiences in CWS during COVID-19 (N=455)

Category	Element	Sentiment Polarity	Review example
Staff (336)	Hospitality, technology support, managing community, operating spaces	(+): 266	*"They take great care and pride in each of their locations which really makes a feeling of belonging."* *"On-site and virtual I.T. support staff to help with any hiccups…"*

Table 2 continued

		(N): 14	*"Everything always works when she's here and makes everyone feel comfy."*
		(-): 56	*"… their customer service is terrible plus there are hidden unreasonable fees."*
Space (270)	Indoor environment quality (IEQ), smell, spaciousness, privacy, outside view, space configurations (i.e., types of spaces, furniture arrangements), atmosphere, size and availability of specific rooms (i.e., conference rooms, phone booths), cleanliness of facilities	(+): 253	*"I have always felt that this place is a safe and tranquil space. The tenants are also respectful and quiet."* *"Great space with all the amenities; full kitchen, gym, rooftop, cafe, and more. Plenty of conference rooms and private phone booths."*
		(N): 5	*"The space has been very functional for regular working, meetings, and even events."*
		(-): 12	*"Facilities are outdated; toilets have black marks all over, water filter has mold…"* *"Our space was loud due to construction, …"*
Service (270)	Networking events for members and non-members, virtual office (mailing, telephone answering), childcare, training, leisure and	(+): 226	*"There are regular happy hours and holiday parties, and there's always a riveting conversation happening somewhere in the building."*
		(N): 34	*"There's also unlimited tea and coffee, a microwave, a fridge, and a toaster oven if you're planning a long study session."*

Table 2 continued

	entertainment programs, free coffee and snacks, pet-friendly policy	(-): 11	*"... There also are no amenities, no coffee, and no snacks..."* *"If you are planning an event here DON'T. They do not have the permit for events..."*
Technology (67)	Internet access (WiFi), printing systems, outlets video conferencing systems, reservation systems	(+): 44	*"... WiFi fast, and the private conference room had all the amenities I needed for the Zoom interviews I had set up."* *"Conference room booking system- Great, easy process for welcoming external guests ..."*
		(N): 21	*"The internet is very fast and ..."* *"Their WiFi kept dropping for six weeks in a row..."*
		(-): 2	*"... you cannot do a video call in them; they take a long time to get your email setup in the system."*
Accessibility (81)	Proximity to home, public transportation, park (nature), and restaurants, easiness to navigate, parking fees, availability, and distance	(+): 61	*"My clients love it because it's right off the freeway, with plenty of parking, and not in the middle of downtown where it can be hectic."* *"Location is ideal as it's close to EVERYTHING from great restaurants to the lake. The lake is a quick 10-minute walk for an escape to meditate or maybe even a picnic lunch."*
		(N): 15	*"The location means lots of food options nearby;..."*
		(-): 5	*"...Parking nearby is tough (and not free unless you walk pretty far)."*

Table 2 continued

			"… the automatic exit clock ahead by 6 minutes, … it is already some of the most expensive parking in the city."
Membership (139)	Affordable fees, additional charges for services, membership cancellation, short-term and long-term contract period	(+): 72	*"They truly are month-to-month and have never increased my monthly rate without sacrificing any customer service."* *"My business and working from home was too distracting, so I searched for a day pass somewhere."*
		(N): 27	*"The rates are very reasonable compared to other options in the area."*
		(-): 40	*"…I'm extremely disappointed they have afforded no pause on memberships (I requested this, and it was just declined)…"* *"… The place nickels and dimes you on everything…"*

* (+): positive, (N): neutral, (-): negative

COVID-19-specific user reviews about CWS management responding to COVID-19 are largely categorized into staff, space, service, and membership (Table 3). In terms of spaces/facilities, installing air-filtration was recognized as a positive reaction by space users. Sanitization of furniture and facilities was indicated as a safety factor. Covid-19 concerns led to additional management by staff to clean and sanitize their spaces as well as to inform and encourage them to wear a mask and maintain social distancing. When CWS managers were not dedicated to keeping the space clean and safe, space users felt unsafe and unsatisfied with the CWSs. From the aspect of an entrepreneur, social distancing regulations required a small enterprise to rent more space and increased the costs for membership. While the staff's responsibilities were increased, COVID-19 required staff to be out of the office or kept the staff from contacting space users. This led to delays in response time for customer requests and consequently, lower satisfaction.

Table 3. COVID-19 related user experiences in CWS

Category	Element	Sentiment Polarity	Review example
Staff (34)	Hygienic responses to COVID-19, cleaning space	(+): 30	*"Staff were extremely friendly and accommodating and were super helpful in explaining their COVID protocols."*
		(N): -	*N/A*
		(-): 4	*"NO SHOW! I got a tour by myself in the middle of a pandemic."*
Space (65)	Office needs during a lockdown, air quality, cleanliness, space configurations for social distancing, amenities (e.g., hand sanitizers)	(+): 52	*"They even installed a brand-new air filtration system."* *"The COVID setup is extremely well done (constant cleaning and use of glass barriers)."*
		(N): 9	*"The spot is very spacious and had plenty of room for social distancing during these times of COVID 19."*
		(-): 4	*"Not sure about the hygienic factor as I haven't seen a change other than stickers being placed on surfaces."*
Service (12)	A sense of belonging, managing members,	(+): 12	*"Honestly, the best part about working here during COVID is … less lonely*

Table 3 continued

	abiding with CDC guidelines		*working out of a coworking space instead of from home...”*
		(N): -	*N/A*
		(-): -	*N/A*
Membership (11)	Changed space use, cancellation, and pause	(+): 5	*“I had a very positive experience when I asked to cancel my membership due to coronavirus. They were very generous to allow me to use my deposit for the last month's rent.”*
		(N): 1	*“… They will only try to return the deposit before the end of the year, using the pandemic as a reason…”*
		(-): 5	*“… We can now only fit half of the people the room claims to hold with the updated guidelines… We pay a lot of money as a small business for this space …”*

* (+): positive, (N): neutral, (-): negative

4.3.2 Descriptive Analysis

The divided datasets, before and during the pandemic, indicated different trends in the reviews; while space was the most mentioned topic before COVID-19, staff was the most mentioned topic during COVID-19 (Figure 8). Notable trends were evidenced in the topics of service and membership. Not surprisingly, the frequency of service-related reviews, including those related to networking and communication events, decreased during COVID-19. Membership-relevant issues were mentioned more during the pandemic than before.

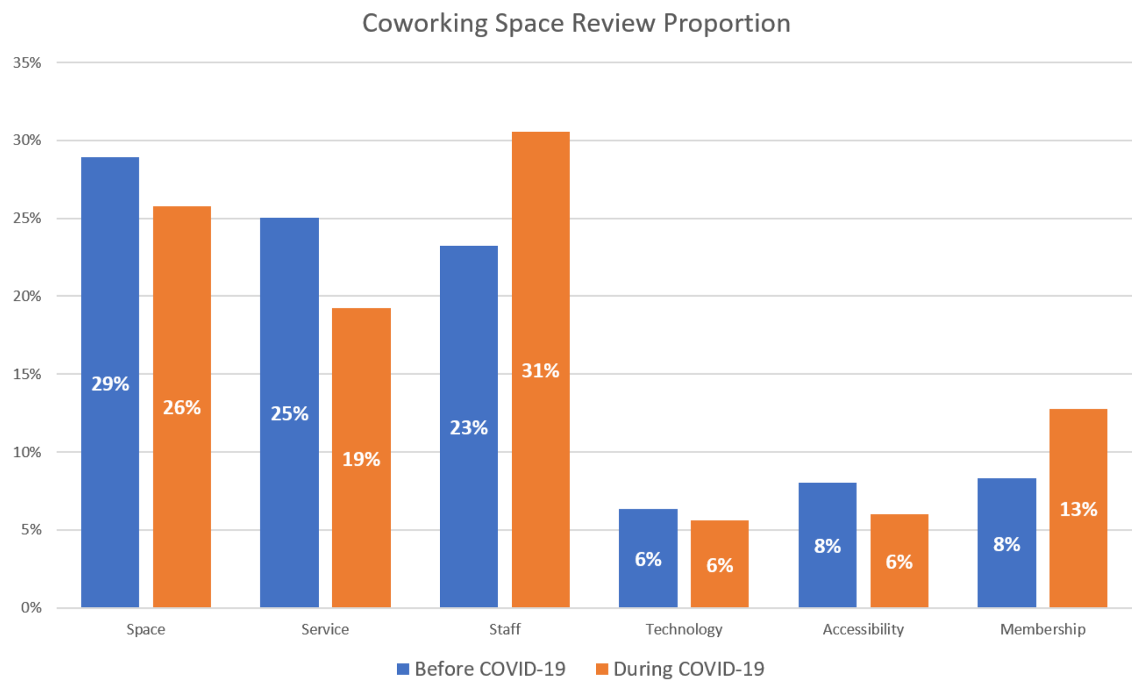

Figure 8. The proportion of CWS reviews about before and during COVID-19

Co-occurring categories are depicted in more detail in Table 4. The categories before COVID-19 showed constant occurrences. For instance, space was the most co-occurring category with others. However, staff became the 1st co-occurring category in terms of membership during the pandemic, while the staff was ranked as the 3rd before the pandemic. Staff was also ranked the 1st co-occurring topic with space. Technology was the 3rd co-occurrence topic with space and service instead of accessibility.

Table 4. Cooccurring category rankings before and during COVID-19

	Before COVID-19				During COVID-19		
	1st	2nd	3rd		1st	2nd	3rd
Space	Service	Staff	Accessibility		Staff	Service	Technology
Service	Space	Staff	Accessibility	➔	Space	Staff	Technology
Staff	Space	Service	Accessibility		Space	Service	Membership
Technology	Space	Service	Staff		Space	Service	Staff
Accessibility	Space	Service	Staff		Space	Staff	Service
Membership	Space	Service	Staff		Staff	Space	Service

4.3.3 Machine Learning: Clustering and Topic Discovery

A total of 12,029 meaningful sentences were selected and reviewed out of 32,783 sentences. The others were not clustered or did not show meaningful attributes. HDBSCAN resulted in 52 clusters. The minimum cluster size was 16 data points among the meaningful clusters. Since clustering is an unsupervised learning method, many clusters consist of meaningless words or phrases with only general and emotional expressions, such as ‘best place,’ ‘great place,’ and ‘highly recommend.’

As a result, 14 clusters were selected to interpret the results. First, refreshment-related reviews (*C1*) were grouped with 1,635 sentences, such as ‘free coffee,’ ‘free beer,’ ‘afternoon snack,’ ‘kitchen area,’ and ‘coffee bar.’ Staff-related reviews (*C2*) were the second-largest cluster that included ‘friendly staff,’ ‘friendly accommodating,’ and ‘customer service’ with 1,352. The rest of clusters were as follows: cleanliness of space (*C3*), including ‘clean space’, ‘clean well’, and ‘place clean’ with 235, size, types, and availability of facilities (*C4*), including ‘private office’, ‘multiple conference’, ‘room large’, and ‘private phone’ with 203, parking convenience (*C5*), including ‘parking lot’, ‘street park’, and ‘plenty parking’ with 192, acoustic comfort (*C6*), including ‘nice quiet’, ‘quiet place’, and ‘comfortable quiet’ with 152, fast internet service (*C7*), including ‘fast internet’, ‘fast reliable’, and ‘WiFi accessible’ with 123, interior design (*C8*), including ‘modern space’, ‘beautiful modern’, and ‘modern décor’ with 94, price and membership related sentences (*C9*), including ‘less expensive’, ‘price reasonable’, and ‘efficient offer’ with 88, outside view (*C10*), such as ‘overlook river’ and ‘outdoor patio’ with 55, visual comfort like lighting (*C11*), such as ‘natural light’, ‘well lit’, and ‘open bright’ with 47, pet friendly policies (*C12*), including ‘dog-friendly’ and ‘service animal’ with 38, community mood (*C13*), such as ‘people nice’, ‘people smile’, and ‘upbeat place’ with 24, and finally, location (*C14*), such as ‘nearby station’, ‘block away’, and ‘Korea-town’ with 16.

The clustering method did not provide additional theoretical insight. However, the results helped to understand the data at a glance with the structured categories by reorganizing the clustering results (Figure 9). The results were regrouped for the illustration according to the main categories. Space included cleanliness, acoustic comfort, interior design, lighting, outside view, and size, types, and availability of facilities; Service involved refreshment, community mood, and pet-friendly policies; Technology included Internet access; Accessibility included parking convenience and CWS locations; Membership was composed of price and contracts.

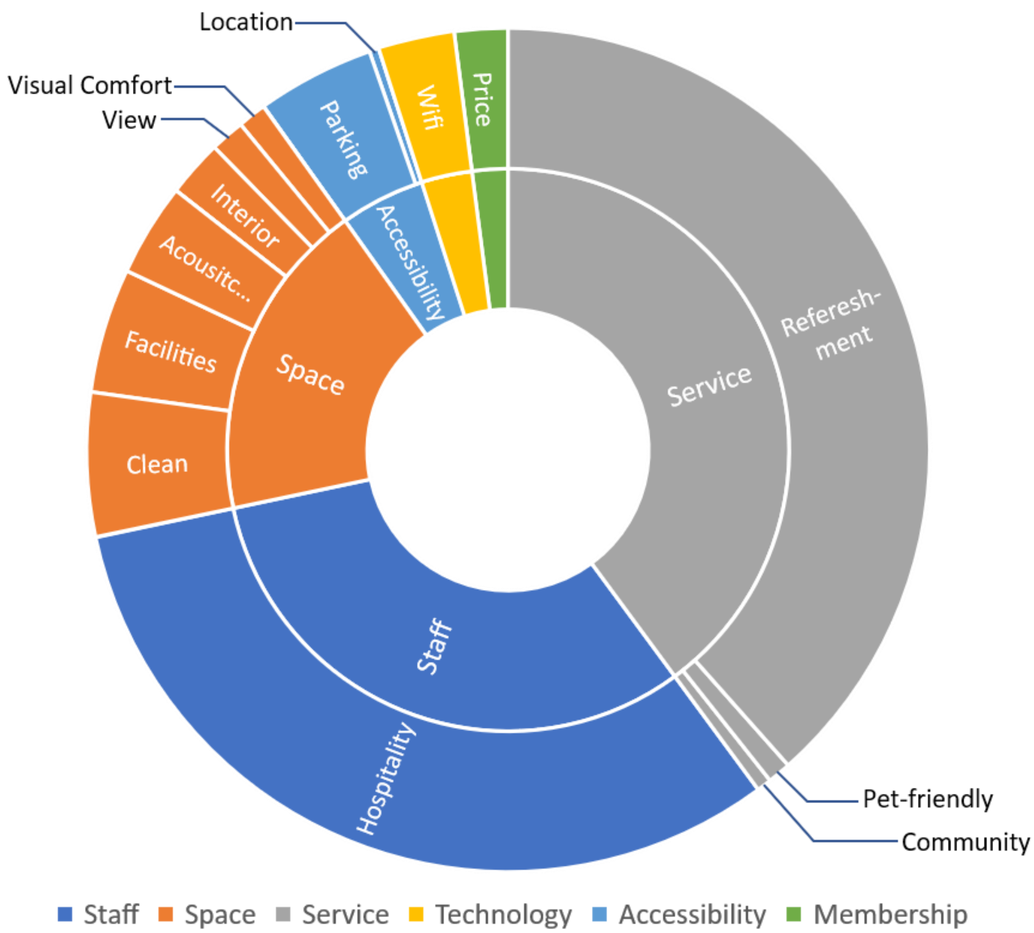

Figure 9. Reorganized clustering results based on the thematic categories

4.3.4 *Machine Learning: Multi-label Text Classification*

The classification models' average overall accuracies and micro average f1 scores were evaluated with the test data set (Table 5). Since SVM, logistic regression, and MLP transformed to a binary relevance in predicting six topics, the training and predicting process repeated six times for each topic. MLP model showed the best performance in the overall accuracy and f1 scores among the four classifiers. The overall accuracy, 85%, was fairly strong in predicting a topic. BERT, the most state-of-the-art model among the classifiers, showed the lowest performance. However, BERT considered the relevance between words, whereas the other models predicted one label at a time.

Table 5. Average overall accuracy and micro average f1 score of predictions or each classifier about the test data set

	BERT	LR	SVM	MLP
Overall accuracy:	74%	81%	84%	85%
F1 score:	67%	73%	77%	78%

The results showed different accuracy in predicting each category from 57% to 89%, as shown in Figure 10 (a). Specifically, in the BERT model, 82% for technology, 79% for accessibility, 78% for membership, 67% for space, 62% for staff, 57% for service; in the logistic regression model, 83% for technology and accessibility, 81% for membership, 75% for service, 74% for staff and space; in the SVM model, 89% for technology, 88% for accessibility, 85% for membership, 77% for space, 74% for staff and service; lastly, in the MLP model, 88% for technology and accessibility, 84% for membership, 79% for space, 76% for service, and 75% for staff. The results of technology showed the highest prediction accuracy in every model. On the other hand, space, service, and staff showed lower performance than technology, accessibility, and membership.

Since a multi-label text classification could have an imbalance issue, f1 scores for all labels were evaluated in Figure 10 (b). The f1 scores of prevalent labels, such as space, service, and staff, ranged from 0.71 to 0.85. However, the models showed lower performance than other labels, such as technology, accessibility, and membership, from 0.21 to 0.68. The models presented an imbalanced data issue to learn. For instance, all the prevalent labels were labeled for a data point, whereas the others were not. The imbalanced data issue led to weak learnability with arbitrarily high accuracy (Schapire 1990). In particular, since the technology category showed invalid f1 scores in all four models, additional 100 data points that included technology-related words were labeled to increase the number of technology labels to resolve the imbalanced data issue. However, BERT still showed invalid f1 scores on technology, accessibility, and membership, which means that the trained BERT model cannot be used to identify CWS reviews about technology, accessibility, and membership. The imbalanced data issue was discussed in detail in the discussion section.

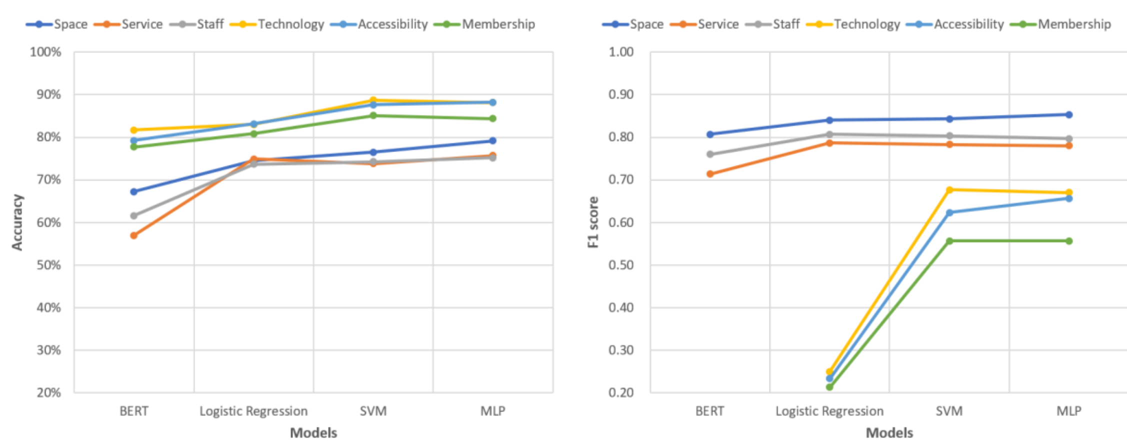

Figure 10. (a) Category level classification performance of prediction accuracy (left), and (b) F1 score (right)

4.4 Discussion

The findings of this study are further discussed in this section, specifically 1) the reasons for dissatisfaction with each category based on the content and descriptive analysis results, 2) changed user preferences in CWSs, 3) the outcomes of the clustering methods for the theoretical benefits of the clustering methods, and 4) the validity of the proposed automatic classifiers.

4.4.1 Dissatisfaction with Physical Environment and Service Categories in CWS

Investigating user complaints in the reviews can be a way to measure minimum quality levels of each topic. For example, when CWS provides more free services, the level of user satisfaction is likely to increase. However, such services are not essential and sustainable to operate CWS (Appel-Meulenbroek et al. 2021). Since some neutral sentiment reviews included both positive and negative aspects, neutral sentiment reviews were also investigated alongside negative sentiment reviews to reveal the reasons for dissatisfaction with the physical environment and service factors.

Staff. Reception and hospitality issues were some of the reasons for the negative values. Specifically, there were membership contract-related issues stemming from the pandemic. Staff's delayed and unclear responses were the dominant causes of dissatisfaction.

Space. IEQ issues were pointed out by some CWS users, such as dark rooms or space, loud spaces, poor hygiene conditions, insufficient social distancing from others. Proximity to other CWS users not only caused discomfort at a workstation but also was related to privacy issues: *"The coworking area (lounge) is packed with people working* *shoulder to shoulder, trying to squeeze every bit, making it not a comfortable place to work."* and *"… I had occasions where there were people visibly interfering with my computer screen in certain ways…"*.

Service. As the clustering analysis showed many reviews about refreshments, the provision of refreshments was considered an essential component in CWS. Free or inexpensive refreshment service was suggested to maintain or increase satisfaction levels. In addition, coffee corners have been highlighted as a facilitator for interactions between coworkers (Spinuzzi 2012). Interestingly, this finding presents that refreshment is a driver that influences both individual users' satisfaction and interactions between coworkers.

Technology. In addition to WiFi stability, technology issues with using CWS resources were found, such as problems with printers, email systems, and room reservation systems. Work-supportive technology became important to increase the quality of CWS. Security concerns also existed like *"Hot Spot Desks at the coworking space are using a shared WiFi system... you are signing into a shared network where anyone with any level of computer skills can possibly hack into your computer and shadow your work"*.

Accessibility. Parking lot location and fees were mentioned with a negative sentiment. Accessibility was found to be one of the most important factors in terms of preferences in CWSs (Appel-Meulenbroek et al. 2021). In the study by Weijs-Perrée et al. (2020), accessibility by car was significant among the user preferences, such as the availability of parking lots. The results of this study also show the importance of the convenience of parking, including availability, the distance from parking lots to CWS, and parking fees. This attribute was important for members and their clients whose feedback was critical for the members.

Membership. Membership contracts are another critical factor for satisfaction related to the users of CWS (Appel-Meulenbroek et al. 2021). Likewise, the results of this study presented that many members complain about hidden fees in contracts. The following paragraphs discuss membership cancellation and suspension issues during COVID-19 while addressing RQ4.

4.4.2 Changes in User Preferences for CWS Factors During the Pandemic

The results in this study showed that staff's roles were expanded to create an environment that felt both safe from infectious diseases and welcoming for a sense of community. The proportion of staff in the reviews increased from 23% to 31%. Co-occurring topics showing the interconnection between different categories in Table 4. Cooccurring category rankings before and during COVID-19 imply that staff became the first ranked topic within the space topic. User experiences indicated that they felt safe and comfortable when staff kept space clean and addressed users' behaviors that did not adhere to Center for Disease Control and Prevention (CDC) guidance. For instance, when staff did not care about mask wearing, social distancing, and the cleanliness of spaces, users felt uncomfortable and unsafe using those spaces. The absence of staff to respond to space users' requests was another factor that negatively influenced satisfaction.

Physical changes in CWS were also related to user satisfaction. New space configurations and equipment for fresh indoor air were mentioned with positive sentiment, such as spacious seat configurations and installation of plastic walls between seats. COVID-19 led to improved indoor air quality and heightened hygienic standards to prevent infectious diseases through constant airflow, frequent workstation cleaning, and air filtration in nonmedical facilities, including offices (Nembhard, Burton, and Cohen 2020; Prabhakar et al. 2020). Some of these physical changes are in conflict with conventional operations in CWS because CWS is designed to maintain proximity and openness to promote knowledge-sharing activities. Thus, new arrangements are desired to meet the changing preferences, such as being able to maintain eye contact but dividing the space with glass or plastic walls or dividers.

Although networking events and interactions between coworkers decreased during COVID-19, the need for a sense of community remained. The results of the content analysis highlighted community. As gig workers and small business owners and employees have used CWSs for working alone but together (Spinuzzi 2012), employees who wanted to avoid loneliness at home started to use CWSs during the lockdown. Many employees who had not previously worked from home were forced to work from home during the pandemic. The perception of working from home has changed, and many employees prefer a hybrid workplace embracing workplace flexibility (Slack 2020; Yang, Kim, and Hong 2021). However, as the review presented, CWS can be an alternative office to resolve professional and social isolation resulting from working from home.

In addition, COVID-19 led to membership issues, including cancellation, suspension, and limited occupancy issues (Table 3). This likely contributed to the increased percentage of membership discussion in reviews from 8% to 13%. Staff was ranked the first co-occurring category regarding membership during the pandemic. Many CWS did not accept the requests, and the requests and complaints were ignored by the staff. This temporarily disorganized customer service led to low ratings and dissatisfaction with CWSs. On the other hand, CWSs that refunded the membership fees in response to requests showed high ratings. Monetary complaints critically influenced overall satisfaction with CWSs notwithstanding the quality of the physical environment.

Lastly, although the percentage of technology in reviews was not changed, technology needs varied after the pandemic. Technology was ranked third with respect to space and service in the co-occurring topic table (Table 4), replacing accessibility. Specifically, the required reliability level of Internet connection was heightened in order to support stable video conferencing which replaced in-person meetings (MacMillan et al. 2021). The availability of spaces and equipment for Zoom meetings became more important together with private phone booths and conference rooms as demonstrated in comments such as *"... WiFi fast, and the private conference room had all the amenities I needed for the Zoom interviews I had set up."* In this aspect, not only conventionally considered technology components, such as stable WiFi, distributed outlets, and printers, but also spatial components associated with technology, such as a large screen, office phone booth, and virtual conferencing room. The technology components and spatial components should be considered together according to space users' activities.

Through the analysis, the combinations of CWS categories with staff coordination were recognized as important for user satisfaction (Figure 11). In particular, the complementary roles of community managers facilitated coworkers' activities through assisting with room reservations and technology use, accommodating networking events, and creating a friendly, community atmosphere for professionals. Community managers create distinguished interactive and collaborative places according to their understanding of the activity as a coworking host (Merkel 2015). Brown (2017) emphasizes community managers' roles that curate the spaces for interactions, including knowledge exchanges and collaboration. In addition to interactions, this study indicates that community managers should accommodate CWS members to get work done and interact with others in virtual environments through offered facilities, technology, and services.

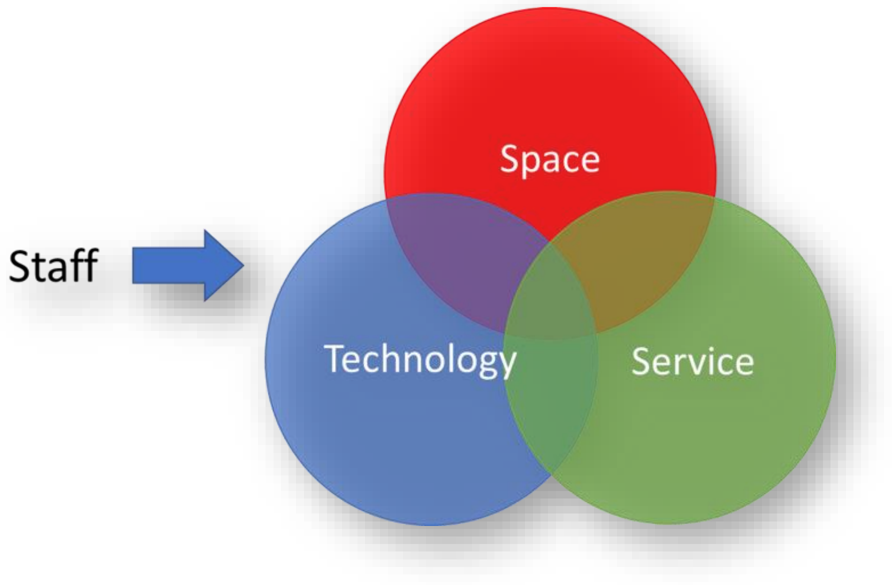

Figure 11. Relationship diagram of CWS experience

Another key responsibility of a community manager is marketing. Community managers introduce the space to new potential users and attract them to join as members. In summary, the roles of a community manager are broad and critical in operating CWS and encompass marketing, managing memberships, coordinating network activities, keeping amenities, and supporting the users' individual activities (Huang 2021). Given their specific roles and responsibilities in the areas mentioned above, fair compensation and treatment of community managers remain to be further examined and improved upon. In fact, Glassdoor, where salaries and reviews of companies are shared publicly by employees, shows that the salaries of community managers vary widely among CWS companies.

4.4.3 Insight from Clustering Outcomes

Parking lots (*C5*) and pet-friendliness (*C12*), factors that were less significant in previous studies, were discovered to be desired factors in the results. First, the central locations of CWS in cities provide convenient accessibility for the CWS users (Bouncken et al. 2020; Waters-Lynch and Potts 2017). However, previous studies have rarely covered parking lot features, although they are closely associated with accessibility with respect to CWS and users' preferences (Weijs-Perrée et al. 2020). The bigram words captured capability, proximity, and parking lots' prices. Interestingly, the results mentioned parking lots more than location by CWS users. These can be a general characteristic of the U.S., where accessibility by a car is important. Second, pet-friendly policies were also mentioned. In this result, the preferences for pet-friendly policies were polarized. Pet-friendly policies can be taken into account by surveying users' preferences in a respective CWS.

This study confirmed the performance of SBERT and the clustering method to observe the overall compositions of the data set. However, unsolicited data consists of the most common opinions, so it is challenging to find an insightful opinion without a specific question (solicited question). Particularly, it was difficult to delineate services that could be attractive to users but are provided by only a few CWS. For instance, a virtual office service was discovered through content analysis, which entrepreneurs preferred. The service provided an office address for members, a reception service for members' clients, and a virtual assistant with phone answering services. Childcare facilities and services were also marked as a favorite attribute for working parents caring for their children during working hours. Therefore, additional search methods should be performed not to miss the valuable insights embedded in a few reviews.

Studies using online reviews of the physical workplace have been conducted by manually selecting words, searching the words in the reviews, and analyzing sequential sentences (Chinazzo 2021; Villeneuve and O'Brien 2020). For instance, when a sentence has a word related to indoor conditions, the previous sentence, the subject sentence including the word, and the next sentence are selected for sentiment analysis. However, other sentences that are not adjacent to the subject sentence can be related. Additionally, such a labor-intensive process – selecting the relevant words to search for in reviews and repeating the process – can be extremely challenging when addressing voluminous data. Therefore, a systemized process is required to group relevant sentences in a review and understand the context with other given information such as ratings.

4.4.4 Model Validation of Classifiers

MLP showed the best performance; the overall accuracy calculated by Equation 3 and Equation 4 was 85%, and the f1 score calculated by Equation 2 was 78%. These results are fairly strong. However, BERT was utilized to consider the correlation between categories (labels) since a binary relevance transforms multi-label classification tasks into independent binary classification tasks (Zhang et al., 2018). The imbalanced data issue was magnified in the BERT model. The BERT model showed biased predictions in the case of certain infrequent cases where none of technology, accessibility, and membership were able to marked among the data points. In other words, marking such cases as 0, which means that the review does not include technology, accessibility, or membership-relevant words or phrases, is more efficient and effective to increase accuracy. This is why the f1 score should be computed to evaluate the models. In this vein, the BERT model was invalid in predicting the infrequent categories. Technical methods to address imbalance issues are suggested for a future study.

Multi-label synthetic minority over-sampling (MLSMOTE) is suggested to resolve the imbalanced label issues for the BERT model and to consider the correlation between the categories (Charte et al. 2015). Essentially, MLSMOTE is one of the methods to oversample, or to increase the number of samples. This method augments infrequent class data points. For example, reviews that only include the technology category are duplicated to increase the number of reviews solely mentioning technology. The MLSMOTE algorithm automatically performs this process. MLSMOTE provides mathematical criteria to determine such minority class and augment the minor class data points using n-nearest neighbors of the reference minority class data point (Charte et al. 2015). The process is repeated to balance the data sets. MLSMOSTE is expected to resolve the imbalance issue to utilize BERT to observe the correlation between the classes.

4.5 Limitations

This study has limitations with data collection centered on CWS in urban and suburban cities close to the urban cities. This study may not cover user experiences and preferences of CWS in cities with different locational and industrial attributes. The demographic information of the users who wrote reviews was not considered in this study. Data collection taking into account gender, occupation, age, and ethnicity may result in different preferences. The differences between online platforms are not considered in this study as well. For instance, Foursquare Swarm users may have different perspectives and opinions in evaluating CWS.

The sentiment analysis showed that the large proportion of reviews is positive. This result is observed in online platforms such as Yelp and AirBnB (Fradkin et al. 2015; Potamias 2012). Potamias (2012) pointed out that the first reviews show bias to overestimate the merchants. In fact, the average of the first reviews was 4.1 stars whereas the average of 20^{th} reviews was 3.69 stars. The average number of reviews was 3.3 reviews per CWS in this study. There might be an overestimation in the reviews.

Although some of the results of this research can be adopted in a corporate office setting, there would be constraints due to the different characteristics between CWS and corporate offices. The priority of CWS is to attract CWS users and increase the number of memberships for profit, whereas a corporate office is usually provided to specific occupants such as employees of the tenants. Services and amenities are very important management tools from the viewpoint of not only work support but also hospitality. This may lead to different approaches depending on the space managers, budgets, and occupants' attributes.

CHAPTER 5. STUDY II: DATA INTEGRATION TO BUILD ROBUST MACHINE LEARNING BASED TEXT CLASSIFICATION MODELS FOR FACILITY MAINTENANCE DATA

5.1 Introduction

Study I presents the general preferences for CWS factors. In Study II, we explore the performance of a specific CWS. Particularly, Study II investigates maintenance issues that influence CWS users' satisfaction such as satisfaction with temperature, lighting, and air quality, and convenience in using CWS facilities including a kitchenette, restrooms, appliances, and electric devices. Mechanical, Electrical, and Plumbing maintenance issues indicate the performance of a CWS management regarding facilities. As we discussed the applicability and usefulness of NLP and machine learning techniques, the data generated from a CWS can be analyzed by using such methods. However, accumulating the amount of data to build a robust machine learning model from an individual CWS may be protracted. Data integration with the data generated from other CWSs and buildings could be a way to resolve this issue. Thus, this study investigates the adaptability of data from other facilities and the applicability of data integration with the single building dataset where a CWS is located. Computerized Maintenance Management System (CMMS) data that stores maintenance requests and records are utilized in this study.

CMMS data have been investigated to discover primary factors and rules (Bortolini and Forcada 2020; Gunay, Shen, and Yang 2019) and automate maintenance request classifications (Hong, Kim, and Yang 2022), tasks, and priority assignments (Mo et al. 2020) utilizing natural language processing techniques and machine learning. Hong et al. (2021) verified the performance of multiclass classifiers with 85% accuracy in 24 HVAC classes and showed the feasibility of performance improvements through analyzing human and input error cases. The previous studies applied natural language processing (NLP) and machine learning algorithms to analyze the unstructured text data in a maintenance dataset.

A critical issue with CMMS is interoperability. In contrast to data management in Architecture, Engineering, and Construction (AEC) where there is a major design software called Autodesk Revit Architecture for computer-aided drafting (CAD) and building information modeling (BIM) (Berwald 2008; Weber and Hedges 2008), CMMS software varies and has different systems to record maintenance logs with diverse metainformation criteria. Fang et al. (2019) suggest that NLP and machine learning can also enhance interoperability by automatically categorizing existing maintenance datasets into standardized coding format, for instance, the Building Cost Information Service (BCIS) code developed by the Royal Institution of Chartered Surveyors (RICS). In fact, buildings comprise essential facilities and equipment and share common maintenance issues, such as heating, ventilation, air-conditioning (HVAC), plumbing, electrical maintenance, and fire safety. In this context, the applications of NLP and machine learning can be considered a means of mitigating interoperability problems.

However, the data in previous research were retrieved from one CMMS software, and the same categories were considered for testing classification performance. This approach cannot handle data insufficiency to build a robust machine learning model. This study addresses the data insufficiency problem by integrating maintenance data retrieved from different CMMS programs by adapting categories in different datasets to targeted categories and automating category inputs using machine learning models. The maintenance issues are called maintenance categories, and the maintenance categories consist of sub-categories of specific tasks. In this context, we can infer two significant conclusions: 1) similar categories are expected in different maintenance data sources, and 2) machine learning models can be trained with the integrated datasets.

Yet, the development of the machine learning models based on the integrated datasets requires a careful adaptation process to identify and fully reflect the differences in maintenance data structure, maintenance of more or less complicated building systems, use and type of buildings, building operation hours, and building scale. For instance, healthcare facilities are most likely to have more maintenance categories to address the higher complexity of their building systems. Specifically, data reviews and selection of different data sources should be carefully conducted for successful data and model integration (Dong and Rekatsinas 2018). In order to build and train machine learning models with integrated maintenance datasets, data adaptation steps are required. The big dataset would include more categories and attributes in the aspects of various building types and large scale. Thus, the different meta information and features should be efficiently adapted through a data manipulation process. Feature engineering is suggested to review and clean such an unorganized and large volume of data (Scott and Matwin 1999); for example, the attributes of the complementary datasets can be efficiently reviewed using feature extraction skills and provide researchers with an initial intuition to determine the fitness to integrate.

Additionally, the amount of data collected from CMMS software can differ according to the number of buildings that a facility management (FM) department or manager oversees. In other words, since with more data to train a machine learning model comes greater confidence it can perform (Brownlee 2018), data from a single or small building may not be sufficient to train a machine learning model. If different maintenance datasets share similarities, larger datasets can contribute to making the machine learning model of single or small building datasets more confident and robust. In addition, the capability of FM teams varies to resolve issues. Small FM teams would have fewer professionals, and maintenance issues that can be internally handled may be smaller than relatively larger FM teams. This would lead to differences in the number of category types. Therefore, the similarities in maintenance datasets from different CMMS and buildings should be considered.

Hong et al. (2021) and Mo et al. (2020) claim that automated classification of maintenance requests enables facility managers and field staff to save time to find a correct category and reduce input errors. In doing so, sufficient data is required to construct a machine learning-based classification model. However, the applicability of data integration of maintenance data sources from different CMMS software for training machine learning models has not been tested yet, although there is potential to expand the usage of machine learning models. Thus, this study investigates the applicability of the integration of different facility maintenance datasets – a single office building maintenance data and campus facility maintenance data that are collected by different facility management teams and maintenance management systems – and develops a machine learning model.

This study addresses two research questions: 1) Is the different maintenance request data adaptable into other maintenance data sets to build a machine learning classifier? and

2) Does a machine learning classification model with the combined data show higher accuracy than the trained model with only a single building dataset?

5.2 Methods

5.2.1 Framework

Figure 12 illustrates the framework of the study to demonstrate the applicability of data integration in machine learning between the two datasets, through which big data from public and open sources becomes valuable to accelerate machine learning applications in small and medium-sized enterprises and institutes. This process aims to increase the accuracy of machine learning models that classify sub-categories in mechanical, electrical, and plumbing (MEP) maintenance issues of a single building. The text classification process was adopted from Hong et al.'s (2020) study. The computational process was programmed in Python. The Pandas and Regular Expression libraries were employed to efficiently find patterns in text descriptions and deal with the maintenance datasets. The natural language toolkit (NLTK) library was utilized for NLP that preprocess the text descriptions before applying them into machine learning models. Multiple machine learning models were tested to select the best algorithm using the Scikit-learn library.

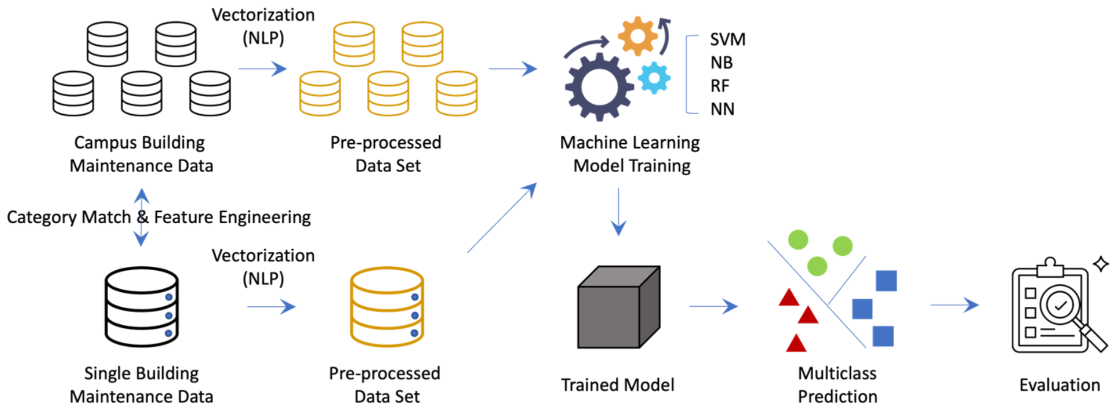

Figure 12. Data integration of different maintenance datasets

5.2.2 Data Collection and Cleaning

This study used two maintenance datasets from different CMMS software. One dataset from academic buildings is retrieved from the Georgia Tech Facility Department, called a *campus dataset*. The campus maintenance data was collected from January 2017 to March 2021. A total of 143,170 data points was collected; 117,173 data points remained after removing duplicates.

Another dataset is retrieved from an office building one block from the Georgia Tech campus, called an *ATDC building dataset*. The ATDC building has been rented out to multiple entities, such as branches of large companies, start-ups, a school department, and a fitness club. The building consists of 20 common offices, collaborative spaces, maker spaces, and gym space. The office maintenance data was collected from January 2019 to June 2021. Each request and the response to the request are described in separate rows in the ATDC dataset. For example, one of the requests was, '*The electrical outlet for the coffee pot does not appear to be working in the* ***Faculty/Staff Break Room #5000. Please***

investigate., Notifications: Engineering Group via email.’, and the response was recorded as ‘Task Status: Completed - Date of arrival is Jan 02, 2019, 10:52 AM; No acknowledgment captured, assumed as Jan 02, 2019, 10:52 AM; Date of completion is Jan 02, 2019, 10:52 AM, Message: ***Replaced GFCI outlet.*** No Notifications Sent’.

The example maintenance issue above shows that the raw data included unnecessary information and descriptions within multiple rows. Thus, the rows were merged into one row, and unnecessary information was removed to improve the classifiers’ performance with essential text data. As a result, 2,596 data points were collected in total. The demonstration of interoperability was conducted over mechanical (HVAC), electrical, and plumbing maintenance issues because those areas are the major functions needed to operate a building and the most frequent issues in maintenance (Atkin and Brooks 2015). Data points, including a problem code (sub-category), were filtered and utilized in the following process.

5.2.3 Category Matches and Feature Engineering

Meta-information (i.e., sub-categories of maintenance categories in this study) were reviewed comparing the datasets of the campus buildings and the single office building called ATDC. The data points in the same sub-categories of the campus buildings were selected and adapted to the maintenance dataset of the ATDC building. The campus and ATDC facilities share major maintenance issues; the list of shared problem codes in electrical, plumbing, and HVAC maintenance issues are highlighted in Table 1. The table indicates that the problem codes of the ATDC building were adopted from the campus buildings’ when the sub-categories were the same. In the other cases, new alphanumeric codes were given to distinguish them from the existing categories, for instance, 05Z1 for 'AFTER HOUR OPERATION REQUEST' that does not exist in the campus building dataset.

In the category of mechanical (HVAC) issues, there was an operation-specific sub-category for the office building: 'After Hour Operation Request' while the issues of 'Too hot' and 'Too cold' matched the campus building dataset. By reviewing the three categories, the matched sub-categories were identified in Table 1. The same sub-category data points in the campus building data set were adapted and merged into the ATDC building dataset.

In relation to electrical maintenance issues, circuit breaker issues were named 'No Power/Reset Breaker' in the campus building dataset and 'Tripped Breaker' in the ATDC building dataset. The request descriptions showed how similar they were; for example, 'The electrical outlet for the coffee pot does not appear to be working in the Faculty/Staff Break Room #5246.'in the ATDC office building and 'IBB-Room 2431 Outlet does not work.' The sub-category 'Light out' was exactly matched. However, due to ambiguity, 'Electric/ Repair' could not be matched with the other categories in the campus building datasets; 'Electric / Repair' only accounts for 31 items out of a total of 1944 in the electrical category. All of them were issues relating to insufficient light or outages.

In the case of plumbing, one sub-category in the campus building dataset was divided into two sub-categories in the ATDC building dataset. For instance, the facility management team in the ATDC building divided clogged drains and pipes into two types: sink and toilet. In this case, toilet-related clogging issues were selectively separated from '04A' when a maintenance request included terms ‘toilet,’ ‘restroom,’ ‘men’s,’ ‘women’s,’ and ‘ladies.’ As a result, ‘04B’ and ‘04C’ were leak issues that were merged into the ‘Leak’ category in the office building data set.

In addition to the domain knowledge-based strategy, similarity of the two datasets was explored by feature engineering. First, a chi-square was utilized for feature selection by testing the independence between the occurrence of a specific word and a category in this study (Shah and Patel 2016). Then, uni-grams and bi-grams regarding each category were selected. This method supports a strategy with facility management knowledge by presenting the data features. This step was performed with preprocessed data to be vectorized and precise.

Table 6. MEP maintenance categories and sub-categories

Campus Buildings, N= # as-is (# adopted)		ATDC Building, N= # as-is (# combined)	
Problem Code (Electrical), N=1,038 (698)		Problem Code (Electrical), N=427 (1125)	
03A	NO POWER/RESET BREAKER●	03A	ELECTRIC/ TRIPPED BREAKER●
03B	RESTORE POWER	03C	ELECTRIC / LIGHT OUT●●
03C	LIGHTS OUT●●	03Z1	ELECTRIC / REPAIR
03D	INSTALL OUTLETS	03Z2	ELECTRIC / MISCELLANEOUS
03E	BATTERY		
03F	ASSIST IN BUCKET TRUCK		
03G	OUTSIDE LIGHTS		
03I	GENERATOR MAINTENANCE		
Problem Code (Plumbing), N=2,734 (1,452)		Problem Code (Plumbing), N= 91 (1,543)	
04A	UNSTOP FIXTURE ▲	04Z1	PLUMBING/CLOGGED SINK ▲
04B	FIXTURE LEAKING/ RUNNING ▲ ▲	04Z2	PLUMBING/ CLOGGED TOILET OR
04C	PIPE LEAK ▲ ▲		URINAL ▲
04D	WATER FOUNTAIN	04Z3	PLUMBING/LEAK ▲ ▲
04E	LEAKING VALVES		
04F	INSTALL SINKS		
04G	PLUMBING REPAIR		
04H	RUN CAMERA INSIDE DRAIN		
04I	POTABLE WATER FILTER		
Problem Code (HVAC), N=4,553 (2,152)		Problem Code (HVAC), N=601 (2,753)	
05A	TOO HOT■	05A	TOO HOT■
05B	TOO COLD■■	05B	TOO COLD■■
05C	AC LEAK	05Z1	AFTER HOUR OPERATION
05D	REPAIR ICE	05Z2	REQUEST
05E	MAKER/REFRIGERATOR		OTHERS
05F	MOVE THERMOSTAT		
…	INSTALL FILTERS		
05Y	…		
	REFRIGERATOR FREON		

Note. Matched sub-categories are marked with symbols (●, ▲, ■)

5.2.4 Processing through NLP

Since maintenance request data consisted of unstructured text descriptions, it was preprocessed and vectorized before the classification step using machine learning models. The preprocessing steps were implemented by lemmatizing, tokenizing, and removing stop-words from the maintenance requests in (Bird, Klein, and Loper 2009). All words were converted to lowercase first before the pre-processing steps. Tokenization: the sentences in data sets were separated into a word level. Lemmatization made words an original form. For example, 'leaked,' 'leaking,' and 'leak' were analyzed as an ATDC and original word, 'leak.' The researcher could make a computer understand that the essential meanings of these words are the same through lemmatization. Meaningless words in the category classifications were removed as stop-words, for instance, 'his,' 'yours,' 'which,' 'that,' and 'are' so that those words will not be considered in the text mining process. In addition, symbols and numbers were removed from the descriptions because numbers and symbols were not meaningful when classifying the types of maintenance work orders such as room numbers and floors.

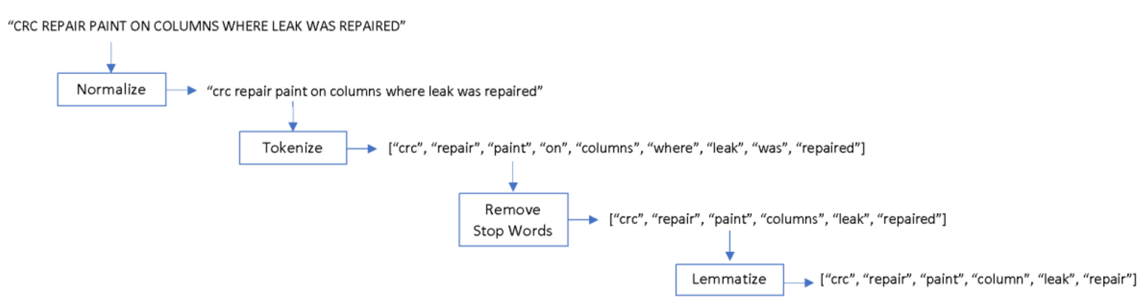

Figure 13. Illustration of pre-processing

After the preprocessing steps, finally, Term Frequency-Inverse Document Frequency (TF-IDF) was utilized to vectorize each text maintenance request. TF-IDF reduces the impact of frequently occurring words that do not have unique meanings (Pranckevičius and Marcinkevičius 2017; Ramos 2003). Uni-gram and bi-grams were used to tokenize which helped interpret the classified results and increase text classification accuracy. That uni- and bi-grams were adopted to find out which combinations make the best performance in predicting a problem code of a work order description. A length of a sequence of individual words is n in n-grams (i.e., unigram and bigram). A maintenance description was split into the n number of continuous words, and the split words were grouped. For instance, regarding a problem code, '05A, Too Hot', 'hot' and 'hot room' are uni- and bi-gram, sequentially. Then, the grouped words were vectorized, and the features were calculated through TF-IDF. The feature of a word or n-gram words were calculated in Equation 1, where it represents the multiplication of the frequency of a word (TF) and inverse document frequency (IDF).

Equation 5. TF-IDF for feature extraction

$$w_d = f_{w,d} * log\left(1 + \frac{|D|}{f_{w,D}}\right)$$

The total number of documents is D, w is a word, and d is a document as a subset of D. TF is the feature of a word frequency, $f_{w,d}$, that the number of a word appears in a document. It is multiplied by IDF that the total number of documents, D, is divided by the frequency of the word in the entire documents, $f_{w,D}$. When IDF is closer to zero, the word means common in the total documents. To avoid an extreme case in IDF, number 1 is added into IDF and logarithm is applied.

5.2.5 Machine Learning based Multi-class Classification Models

Text classification is one of applications of machine learning methods that categorize documents into organized classes. There have been many efforts to utilize a machine learning-based model for solving a text classification problem. For example, Joachims (1998) considered Support Vector Machine (SVM) as the most appropriate machine learning model in a text classification problem. Pranckevičius and Marcinkevičius (2017) used five different machine learning algorithms to perform multiple text classification of the Amazon product review data. They identified that the logistic regression model scored the highest accuracy compared to the other models. In this study, the preprocessed data set was plugged into machine learning models to classify each maintenance request. Four machine learning algorithms were utilized: Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Random Forest (RF), and Multinomial Naïve Bayes (NB).

First, SVM is one of the most popular classification methods. It finds a hyperplane that maximizes the margin between data points that have different classes. The more margin, the more confidence to classify a new data point (Cortes and Vapnik 1995). SVM also shows a good performance in text classification specifically of a short text document (Joachims 1998). By applying kernel, SVM can classify not only a linearly separable data set, but also non-linearly scattered data points. Tuning hyper parameters increase the performance of a SVM classification model. Second, Multi-Layer Perceptron (MLP) performs a non-linear function through each perceptron with input variables and output variables to achieve targeted values (Rumelhart, Widrow, and Lehr 1994). The weight parameters are updated through an iterative training process to minimize the error between target values and predicted outputs (Bishop 2006). Third, Random Forest (RF) is a supervised learning method used for classification with multiple individual decision trees (Ho 1995). The individual trees are trained in parallel and the majority of decision is used as a final class. Fourth, Naïve Bayes classifier is a simple probabilistic classifier that apply Bayes' theorem assuming that each event is independent (Rish 2001). The model is commonly utilized for a text classification problem because of less computation and easy prediction of a class (Friedman, Hastie, and Tibshirani 2001).

The best machine learning model that showed the highest accuracy was selected among those four algorithms. When training a supervised machine learning model, if more data are used to train the model, higher confidence and accuracy are present (Brownlee 2018). Reviewing the two datasets confirmed that the data patterns and types were similar. In this case, evaluating a classification model using more data points and classes is relatively more reliable than using fewer data points and classes. Therefore, the HVAC dataset of the campus buildings with 25 different sub-categories was used to identify the best performance model; 4,553 HVAC datasets were split into a training dataset (90%) and a test dataset (10%).

The training dataset was utilized to build the machine learning models, whereas the test dataset was utilized to evaluate the performance of the models as unseen datasets. Data points in each category were proportionally split (stratified) to avoid biased model training. Thus, the test data set has ten percent of data points of each category in this study; for instance, if we have one hundred labeled data points as 60 dogs and 40 cats, the stratified sampling creates the test set of 6 dogs and 4 cats in ten percent of the total data points.

In addition, the training set was divided into five considering the stratification of sub-categories for cross-validation. Eighty percent of the training set was utilized to build a model, and 20% of the training set was used to evaluate performance. This step was repeated five-fold. The results of the cross-validation were averaged, and one model was selected. The fine-tuning of hyperparameters was considered if available for the selected model. The selected best classification algorithm was applied in training the classification models of HVAC, Electrical, and Plumbing datasets of the ATDC building and the combined building data sets regarding the ATDC building categories. The cross-validation was also performed of the newly generated classifiers.

5.2.6 Evaluation

The performance of each classification model was evaluated by accuracy and f1 score, calculated based on a confusion matrix (Table 2, Equations (1) and (2)). This matrix presents the instances of predictions and actual values, such as true or false. For example, correct predictions have two cases, such as 1) true positive, meaning that predictions indicate true and actual values are also true, and 2) true negative, meaning that predictions indicate false and actual values are also false. On the other hand, unmatched cases between actual classes and predictions include false positives and false negatives. The best text classification model was selected through evaluation. Then, the selected model was used to compare the results between the single ATDC building dataset and the combined dataset of the campus and ATDC building.

Table 7. Confusion matrix

		Actual: Positive	Actual: Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
Predicted	Negative	False Negative (FN)	True Negative (TN)

Equation 1. Accuracy

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)}$$

Equation 6. F1 Score

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$where\ Precision = \frac{TP}{TP + FP}\ and\ Recall = \frac{TP}{TP + FN}$$

In addition to quantitative evaluations, context analyses of the results were conducted. For example, Hong et al. (2020) identified user category errors and confusing categories. The content of error cases was reviewed and compared with the predicted and actual categories. This context analysis process provided a chance to detect human error and to improve the performance of the text classification models.

5.3 Results

5.3.1 Feature Engineering

The chi-square feature selection results are shown in Table 3. Each category includes five top related uni-grams and bi-grams. Bigrams provide more context rather than unigrams. The feature engineering results indicate the feasibility of integrating maintenance datasets from different CMMS software. First, the results of the ATDC building dataset included specific tenants' names and facility staff names assigned to handle the maintenance-related issues. For instance, 'Gym' was highly related to plumbing issues, and 'Joseph,' and 'James,' who handled these maintenance issues, appeared many times.

The campus building data showed a dependency of maintenance-related terms to each maintenance category. As the single building results included the tenant, specific building names where a certain maintenance issue frequently occurred were presented as the dependency to that maintenance issue. Second, through this feature engineering, the similarity of plumbing maintenance request descriptions between the campus buildings and the ATDC building was demonstrated. The selectively extracted plumbing data points like 'Clogged Toilet and Urinal' and the merged categories like 'Feature Leaking/Running' and 'Pipe Leak' indicated reasonable compositions. These feature engineering results supported initial assessments of the feasibility of integrating the different source datasets before training machine learning classifiers.

Table 8. Chi-square feature selection of maintenance requests

Electric		**Plumbing**		**HVAC**	
Campus to ATDC	**ATDC**	**Campus to ATDC**	**ATDC**	**Campus to ATDC**	**ATDC**
03A, No Power/Reset Breaker: *Unigrams:* Reset, trip, outlet, power, breaker *Bigrams:* power outage, breaker trip, power room, outlet work, trip breaker **03C, Lights Out** *Unigrams:* Install, breaker, power, outlet, light *Bigrams:* electrical outlet, room power, light bulb, outlet work, light floor	**03A, Tripped Breaker:** *Unigrams:* Peace, ankobia, outlet, trip, breaker, *Bigrams:* Assign joseph, joseph katz, break room, power outlet, trip breaker **03C, Light Out:** *Unigrams:* Work, power, trip, outlet, breaker *Bigrams:* light floor, bad ballast, floor light, power outlet, trip breaker **03Z1, Repair:** *Unigrams:* Locate, ceiling, possible, address, dim *Bigrams:*	**04A, Unstop Fixture (Clogged Sink):** *Unigrams:* Disposal, sink, kitchen, unclog, drain *Bigrams:* Sink clog, mse room, clog drain, drain clog, sink drain **04A, Unstop Fixture (Clogged Toilet or Urinal):** *Unigrams:* Flush, men, urinal, clog, leak *Bigrams:* Urinal drain, men room, floor men, clog toilet, toilet clog **04B, Fixture Leaking /Running &**	**04Z1, Clogged Sink:** *Unigrams:* Toilet, clog, breakroom, sink, drain *Bigrams:* Gym assign, water leak, joseph katz, assign joseph, room assign, **04Z2, Clogged Toilet or Urinal:** *Unigrams:* Sink, stall, urinal, clogged, toilet *Bigrams:* floor men, gym assign, water leak, woman restroom, hot water **04Z3, Leak:** *Unigrams:*	**05A, Too Hot** *Unigrams:* Room, leak, cold, warm, hot *Bigrams:* Ac work, room cold, hot room, room warm, room hot **05B, Too Cold** *Unigrams:* Project, recalibration, heat, hot, cold *Bigrams:* mse cold, cold air, suite cold, cold room, room cold	**05A, Too Hot** *Unigrams*: Cold, cool, pm, low, hot *Bigrams:* Please cool, please low, low temp, hot degree, hvac hot **05B, Too Cold** *Unigrams:* Pm, temp, normal, increase, cold *Bigrams:* Cold tsrb, increase temp, cold degree, please increase, HVAC cold **05Z1, After Hour Operations Request** *Unigrams:* Schedule, request, operation, hour, pm

Table 8 continued

	Light floor, conference room, freight elevator, light flicker, light outside **03Z2, Miscellaneous:** *Unigrams:* Next, issue, power, panel, leave *Bigrams:* light floor, lobby assign, bad ballast, power outlet, floor light	**04C, Pipe Leak:** *Unigrams:* Drain, run, faucet, clog, leak *Bigrams:* Sink leak, clog toilet, sink clog, toilet leak, toilet clog	Water, suite, drain, leak, clog *Bigrams:* assign randall, randall elizondo, hot water, gym assign, water leak		*Bigrams:* Tsrb floor, hour operation, operation request, HVAC hour, pm pm **05Z2, Others** *Unigrams:* James, replace, Katz, turn, assign *Bigrams:* Joseph katz, james assign, assign randall, randall elizondo, please turn

5.3.2 Model Validation

The cross-validation results from each model were aggregated to select the best performance in machine learning algorithms, as seen in Table 4. SVM showed the best performance than the other classifiers. The radial basis function (RBF) kernel was utilized to increase the performance of SVM with hyperparameter tuning. The kernel function helped classify data points not linearly separable in the original space. The two hyperparameters, penalty parameter (C) and RBF-related parameter (r), were selected 100 and 0.01, respectively. The final SVM text classifier of the HVAC dataset showed 85% accuracy.

Table 9. Cross-validation results of the four classifiers (campus building)

	Data 1	Data 2	Data 3	Data 4	Data 5	Average
RF	0.48	0.49	0.50	0.50	0.43	0.48
NB	0.70	0.73	0.69	0.73	0.73	0.72
MLP	0.81	0.83	0.77	0.82	0.82	0.81
SVM	0.81	0.84	0.78	0.83	0.83	0.82

The tuned SVM model was utilized to generate the text classification models for the HVAC, electrical, and plumbing data of both the ATDC building dataset and the combined (campus and ATDC) building dataset. Table 5 illustrates the aggregated results of accuracy and the f1 scores. The average accuracy of the cross-validation in HVAC and Plumbing indicated a 6% and 19% difference, respectively, from the test results due to insufficient data points to train the machine learning models. Therefore, the ATDC dataset results presented lower confidence than the combined dataset. The classifiers of the combined datasets showed stable accuracies and f1 scores of Electrical and Plumbing between the validation and test sets. The HVAC text classification model presented a 7% difference in accuracy and a 6% difference in f1 score, although the overall performance increased by 8% and 9% in the accuracy and f1 score, respectively.

Table 10. Comparison of accuracy/ f1 score between the ATDC and the combined datasets

	ATDC		Campus + ATDC	
	Validation (Average)	Test	Validation (Average)	Test
HVAC	0.76 / 0.76	0.82 / 0.82	0.84 / 0.85	0.91 / 0.91
Electrical	0.87 / 0.85	0.86 / 0.86	0.93 / 0.92	0.93 / 0.92
Plumbing	0.71 / 0.71	0.90 / 0.90	0.91 / 0.91	0.93 / 0.93

The results demonstrated that complementing the associated data points into the small dataset increased prediction performance. For instance, 'Tenant in suite 2090 has lights out. Not sure how many. Please replace.' was incorrectly predicted as 'Electric/ Miscellaneous (03Z2)' when the text classification model was trained only with the ATDC dataset. However, it was corrected after training the model with the combined dataset.

The category including arbitrary maintenance issues still caused a reduction in the prediction performances in 'Electrical' and 'HVAC.' Specifically, the confusion matrices of the HVAC cases presented the details in Figure 2. The category 'Others (05Z2)' negatively contributed to the performance. In the 'Others (05Z2)' category, only 57% (21 out of 37 cases) were correctly predicted. The detailed results were investigated to identify whether the arbitrary category included the maintenance issues related to 'Too hot' and 'Too cold'; as a result, 14 out of 16 misclassified cases of 'Others (05Z2)' were 'Too hot' or 'Too cold' cases.

HVAC

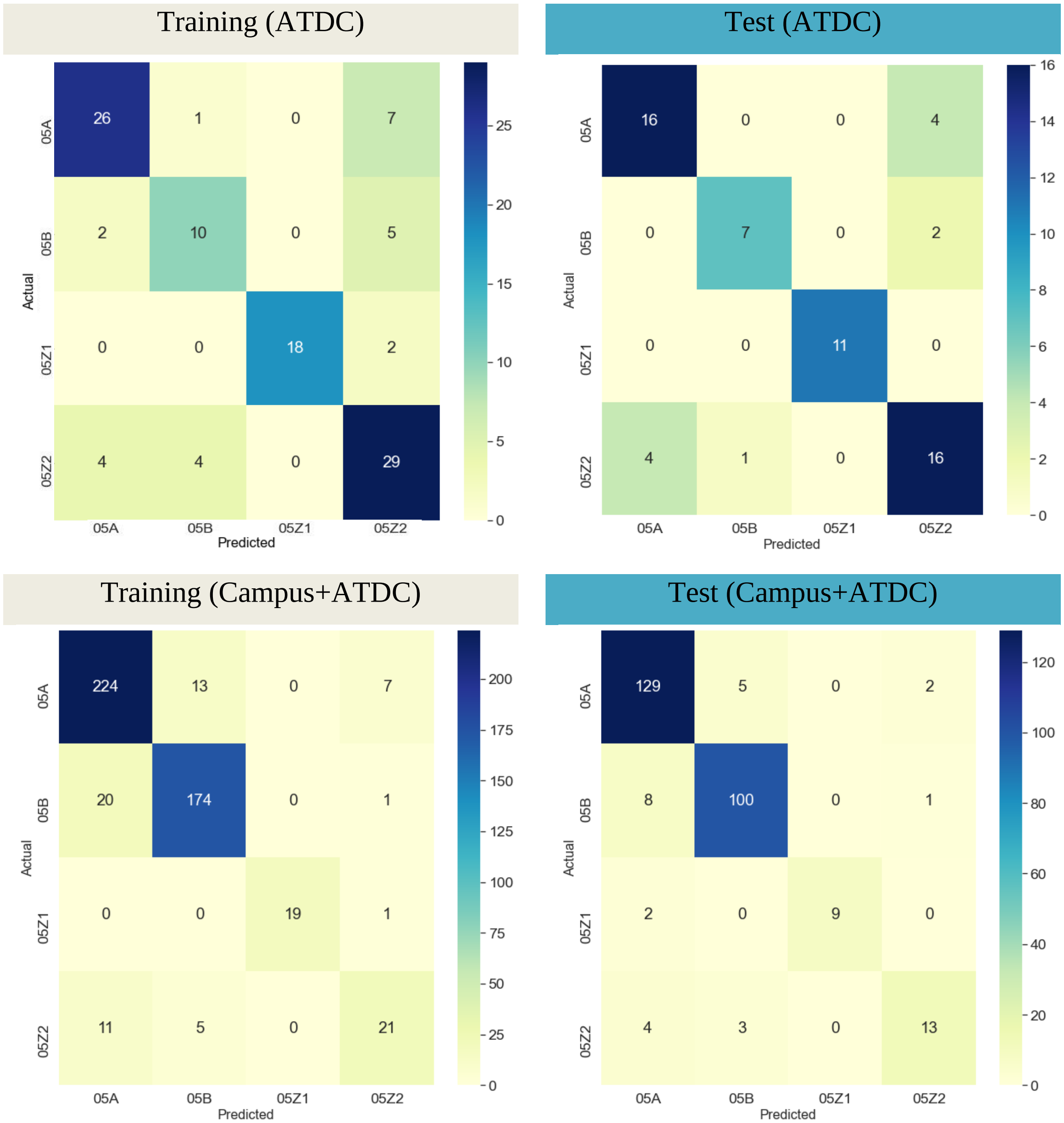

Figure 14. Confusion matrices of the text classifications in HVAC

In addition to the directly merged campus data to the ATDC data, the selectively divided and merged data points, such as 'Clogged sink (04Z1)' and 'Clogged toilet or urinal (04Z2)' in 'Plumbing,' highly contributed to the robustness of the text classification model.

The major error cases in plumbing were ‘Leak (04Z3),’ predicted as ‘Clogged toilet or urinal (04Z2)’. In the validation results of the combined data set (Appendix 3), five cases (55.6%) out of the nine cases were human errors. For example, ‘Urinal overflowing 3rd floor men’s room’ was predicted as ‘Clogged toilet or urinal (04Z2)’, whereas the actual label was ‘Leak (04Z3)’. This type of erroneous input by users was frequently observed in all categories in terms of categorical inputs (meta information).

5.4 Discussion

This study examined the interoperability of facility maintenance data between different CMMS software through feature engineering and machine learning-based text classifications. Integrating domain knowledge and data manipulation techniques reduced the work of manually identifying similarity and categorizing data when adopting another CMMS dataset with different categories. Datasets from the campus-level facility management department included more categories because the department was capable of fixing more complicated and diverse maintenance issues than the facility management department of the ATDC building. In addition to the domain knowledge-based data integration, the text classification experiments of the combined datasets demonstrated the interoperability between the datasets from the different CMMS software. This study demonstrated a novel approach to applying NLP and machine learning techniques in facility maintenance, which accelerates smart and automated facility management in single or small buildings. The implications of accumulated public maintenance datasets were also discussed such as offering a benchmark of maintenance frequencies for retrofits and reorganizing maintenance categories for efficient data management.

5.4.1 *Feature Engineering and Adaptability*

The interoperability of facility maintenance data has been discussed and mentioned by many scholars (Araszkiewicz 2017); data interoperability is important to quickly understand problematic situations and respond in a timely manner to mitigate financial losses. One of the suggested ways to overcome this was to apply machine learning methods, for instance, using text classifications to generate preferred information format (Fang et al., 2019).

This research demonstrated that the maintenance data had major overlapping issues in HVAC, plumbing, and electrical, and they were adaptable to construct robust text classification models for a small dataset. Feature engineering was performed in this study identifying the attributes of maintenance categories and descriptions. For instance, the ATDC building dataset had fewer types of sub-categories in the MEP categories and different sub-categories such as ‘After hour operation requests.’ In terms of plumbing maintenance, clogged issues managed as one sub-category in the campus buildings were separately managed in the ATDC building, such as ‘Clogged sink’ and ‘Clogged toilet and urinal.’ In this case, domain knowledge of facility management and regular expression techniques helped select keywords and extracted clogged toilet and urinal issues from the campus building dataset. The high prediction accuracy of the classifier demonstrated that the specific data points were successfully merged into the ‘Clogged toilet and urinal’ category in plumbing. This result of domain knowledge-based feature engineering is paralleled with the other study results of integrating domain knowledge and feature engineering (Berrar, Lopes, and Dubitzky 2019; Przybyszewski et al. 2017). Critical attributes well-known in those fields were manually selected by modelers to train machine learning models and, in turn, increased prediction accuracy.

In addition to feature engineering for category matches, feature extractions through preprocessing steps enabled the use of unorganized text data formats. Unnecessary system messages and unmeaningful words in the two datasets were removed to improve machine learning classification performance. In this study, removing numbers, orders, and symbols, such as '1st, 2nd, 3rd, [], (), !, etc.,' increased the prediction performance of the machine learning models. On the contrary, specific words in the text, such as building names and room names, positively influenced accuracy. For instance, the chi-square feature selection (Table 3) showed that 'conference room' was highly related to electric maintenance issues and 'gym' presented in plumbing-related issues.

The meta-information, which is tenants and facility staff in this study, contributes to increased classification performance (Mo et al. 2020). At the same time, the meta-information embedded in different larger datasets as references should be cautiously adopted. For example, campus building names were not important to classify an ATDC maintenance issue. Therefore, building names not in relation to the subject building and the dataset should be removed to prevent a negative impact on machine learning classifiers regarding the subject building. In this context, data integration requires identifying accurate and relevant data and cleaning the data according to the targeted dataset (Dong and Rekatsinas 2018). Therefore, feature engineering is critical to achieving the adaptability of datasets from different CMMS software.

5.4.2 *Improved Classification Performance with Difference Datasets*

Integrating the ATDC dataset with the campus dataset increased prediction accuracy quantitatively and improved classification performance qualitatively. The results showed that previously misclassified cases with only the ATDC dataset were corrected after combining the larger dataset and retraining the machine learning model. These findings indicate the potential for improvement in data management in relation to facilities. This study implies that facility maintenance data are worth sharing and that shared data could accelerate achieving smart facility management in single and small-sized buildings.

Accumulating data from public facilities such as government buildings, schools, libraries, and community centers can be utilized to complement data insufficiency to build machine learning models for small-sized or single facilities. The U.S. General Services Administration (2022) utilizes National CMMS (NCMMS) to store all operation and maintenance related records in one database. A report published by GSA (2022) indicates that the NCMMS data is utilized to analyze the types and cost of maintenance of entire government buildings. Sharing this data can significantly contribute to the establishment of maintenance standards and creating data-driven decision-making models by adopting machine learning and NLP techniques. Although the data of private sectors in such facilities, including individual offices and confidential areas, may not be included, data from public sectors, such as classrooms, restrooms, reading rooms, and lounges, will be still valuable because the purpose of data accumulation is to create a big database of general facility management information.

The ENERGY STAR Portfolio Manager could be a good example of utilizing big data collected while managing buildings, providing a benchmark for similar size, type, and area for a building's energy use (Gliedt and Hoicka 2015). Public data with categorical information could offer a benchmark of maintenance frequency and types according to building type, size, and occupant type similar to the ENERGY STAR Portfolio Manager. Therefore, as sensor data integration has been studied to exploit high volume data (Balakrishna, Thirumaran, and Solanki 2020), the integration of user-created data should also be taken into account to promote efficient facility management.

This study also demonstrated that machine learning models were suitable to reorganize data according to updated categories, which is challenging and demanding when manually performed. When data are integrated and adapted, annotation of meta-information based on the targeted data is important to easily access and analyze the integrated data. The results of this study showed successful integration of different data sources and made machine learning classifiers more reliable.

Robust machine learning classifiers can provide facility managers with agility and flexibility to utilize and analyze data. When management strategies and categorical criteria are updated, automated text classification techniques allow facility managers to utilize historical data by relabeling the categories. Thus, the classifiers in this study can assist facility managers in focusing on improving management strategies rather than spending time on technically handling data (Fang et al. 2019). This automated process also can prevent human error with manual annotations (Hong et al. 2022). The labeled data of the campus buildings and the ATDC buildings included human errors, and the machine learning classifiers made correct predictions. In sum, adopting the classification models supports facility managers in concentrating on more critical jobs with higher accuracy.

5.4.3 Implications

Applying data integration and machine learning techniques has two important implications. First, multiple building datasets can provide a benchmark of the frequency of reactive maintenance. A maintenance frequency per square foot is used to compare the condition of facilities (Bortolini and Forcada 2020). In this study, the campus building dataset included more than 100 buildings with similar facility attributes to the ATDC building. The facility conditions of several campus buildings can be utilized as benchmarks to diagnose the ATDC building. This analysis can provide an objective indicator for the building owner and facility managers to establish retrofitting plans. Therefore, the further applications of this study are expected to support data-driven decision-making.

Additionally, the findings in this study offer an opportunity to reorganize the maintenance categories. Ambiguous categories decreased the prediction accuracy of the classifiers. Such categories are also difficult to manually and correctly categorize by experts; junior facility managers' categorization accuracy was below 66% in Fang et al.'s (2019) study. Since maintenance data are inputted by multiple users, such as facility managers and field facility staff, accuracy and intuitiveness of categories are necessary to remain the utility and value of meta-information. In this vein, instances of misclassification help to identify categories that frequently cause erroneous inputs (Hong et al. 2022).

5.5 Limitations

Though the implications of this study are promising, several limitations exist in this study. First, since this case study was performed with two different maintenance datasets, experiments of data integration with more than two datasets can be conducted to demonstrate the validity with multiple data sources. The multiple data sources would have various formats, so multiple feature engineering methods, including the methods used in this study, might be explored to properly manipulate data and extract essential information. Second, although the results showed high performance by combining datasets, the similarity between the two datasets is limited due to different building types. Data from office building stocks could be more suitable for data integration. Finally, although SVM showed a high performance in this study, more state-of-art techniques such as deep learning-based transformers could be tested for improved prediction performance.

CHAPTER 6. DISCUSSION

The two main studies were conducted to provide CWS strategies during COVID-19 by identifying user preferences for facilities and services factors in CWS and changed preferences regarding these factors. The ways in which to apply NLP and machine learning techniques in the physical workplace management research were explored using to social media data and maintenance data. The research validated the applicability of the computational analysis techniques and the data integration.

6.1 User Preferences of Facilities and Services during COVID-19

This research provided six predefined categories arising from comprehensive literature reviews of CWS. Yelp data in relation to CWS were collected and analyzed based on the six categories through content analysis and descriptive analysis and the results from the applications of NLP and machine learning techniques.

Table 11 summarizes user preferences for CWS factors during COVID-19. The findings indicate heightened and detailed management expectations to maintain satisfaction levels of CWS users. It is implied that the changed or added preferences derive from the pandemic and the responses to it such as the demands for remote work by knowledge workers.

First, although a CWS is a private space as a membership-based business, CWS have the characteristics of public spaces that gather diverse and anonymous people and encourage interaction among them. From this point of view, space configurations and occupants' behaviors in CWS must meet the criteria of public space provided by CDC. In fact, the preferences for CWS factors show similar patterns in CWS use as other public spaces. The density of people in the space decreased, wearing masks was required, and in-person interactions were reduced (Jasiński 2020).

Second, as remote work becomes the new normal due to the increased need for hybrid workplaces, the increased frequency and importance of virtual meetings require people to be equipped with a minimum Internet speed to deliver HD video quality for group meetings (Faculty of Arts and Sciences Columbsia University 2022). The changed preferences point to facilities for stable virtual video conferencing such as a high-speed internet service, web cameras, and private space.

Third, the most challenging factor is the community atmosphere in this constrained situation. Interactions between others and colleagues are critical for a sense of belonging (Bartels et al. 2010). Impromptu interactions hardly occur with anonymous people in CWS and community managers should act as agents who interactively communicate among individual CWS users (Brown 2017). This research showed that community managers can introduce a new member to existing members and encourage them to feel included in the community.

The discussions above converge on the most critical factor: the staff (community managers). The hospitality of staff (community managers) is considered the key to adapting CWS to changed preferences and accommodating CWS users in a satisfactory way.

Table 11. CWS management strategies during COVID-19

	Critical factors during COVID-19	Changed or added factors during COVID-19
Staff (community managers)	- Hospitality (e.g. rapid response time) - Hosting networking events and promoting interactions	- Hospitality (e.g. welcoming atmosphere creating a sense of belonging) - Hygiene concerns regarding CWS users and space (e.g. maintaining mask wearing, and social distance between users, and clean space)
Space	- Comfortable IEQ components, specifically, proximity to each other due to privacy	- Space configurations for distance between users, but with eye contact enabled - Equipment and operations for fresh air and a clean space - Rooms and booths for virtual conferencing
Service	- Networking events promoting interactions - Free or inexpensive refreshment and coffee	- Community atmosphere promoting a sense of belonging and reducing social and professional isolation
Technology	- WiFi stability, work-supportive technology (e.g. printers, email systems, and room reservations), and internet security	- High speed Internet for stable virtual communications - Amenities for virtual conferencing (e.g. a large screen, cameras, microphones, etc.)
Accessibility	- Parking lot locations and fees	- N/A
Membership	- Transparent membership fees	- Clear membership policy (e.g. regarding cancellation and space occupancy rates)

This study's would are more applicable when COVID-19 becomes endemic. Because of high rates of immunization in the US, eventually COVID-19 should not be primary as it has been to work from home and interrupting children's school attendance. Instead, it could become seasonal similar to flu (Stieg 2021). In addition to a potential reduced fatality of the virus(es), individual and general public health take attention to our lives, which should lead to preventive behaviors as with other infectious diseases (Feldscher 2021). In order to prevent a fast infection and reduce the impacts, the demand for hygienic responses would remain such as wearing masks and keeping a social distance when getting such disease. Remote work and hybrid work arrangements are becoming permanent (Rubinstein and Hong 2022; Yang et al. 2021). Therefore, virtual communication technology and facilities would become essential to operating CWS.

6.2 Applications of NLP and Machine Learning Techniques

This research explored multiple approaches and utilized NLP and machine learning techniques in the physical workplace management research as shown in Table 12. In the course of the applications, the domain knowledge in CWS and facilities was essential in planning the applications of such techniques and selecting proper algorithms.

In Study I, the thematic categories of CWS factors provided the criteria to organize the unsolicited and unstructured social media data. The text classifications were performed based on the thematic categories. The reorganization of the text clustering enabled the researcher to collect the separate results into a meaningfully combined outcome and provided a clear insight into understanding the massive social media data regarding CWS. Text classification and clustering shed light on how the corporate real estate industry utilizes online user data. Content analysis of massive online data may not be feasible to quickly respond to ever-changing business circumstances. However, machine learning and NLP approaches reduce the time to extract critical features and provide overall insight into the entirety of the data. The results of this study may accelerate an objective and quantitative analysis of the physical workplace.

In addition, the content analysis helped to identify the limitations of the applications of NLP and machine learning. For example, meaningful information in the massive data, though a small portion of data points among the entire data, may be ignored. For instance, childcare facilities and related services contributed to satisfaction. In the machine learning process, it was not caught. This finding suggested that the data analysis should be combined with the relevant knowledge so that researchers can perform a triangulated data analysis.

In Study II, feature extraction, feature engineering, and NLP enabled validation of the adaptability of a maintenance dataset from different CMMS software to a small maintenance dataset to build a confident machine learning model.

This approach is not only applicable in facility maintenance service, but also in CWS service. As Study I showed the categorical information, many CWS have similar facility and service complaints and issues. The data collected from CWS branches can be combined and utilized for training machine learning models. Particularly, small size CWS would take time to accumulate sufficient data to construct a robust machine learning model. However, when the small CWS share their complaint and issue data, they can adopt such machine learning models according to their purposes and criteria.

Table 12. Summary of the applications of NLP and machine learning techniques

	Study I: User Preferences in CWS	Study II: Data Integration
NLP	Preprocessing Feature Extraction (TF-IDF, Word-Embedding)	Preprocessing Feature Extraction (TF-IDF, Chi-square) Feature Engineering
Machine Learning	Text Classification - BERT, SVM, LR, MLP Text Clustering - SBERT and HDBSCAN	Text Classification - SVM, MLP, RF, NB

Domain Knowledge	Thematic Categorization of CWS Factors Content Analysis	Given Facility Maintenance Information - Category and Sub-category Codes Maintenance Types of MEP

6.3 Validation by Professionals

To validate the approaches and the results in this research, the author interviewed three professionals: a community manager of a CWS in midtown Atlanta, GA, a facility manager in the facility management department of the ATDC building, and a facility manager in the Georgia Tech Facility Management (GTFM) department.

First, the community manager said that her CWS had to reconfigure furniture layouts and place sanitizers on every table to respond to COVID-19. The most challenging issue during the pandemic was to attract and retain CWS members. This challenge continued as of the time of the study, March 2022, when facilities had re-opened, and people were returning to normal life. In other words, membership attraction and retention have been the first priority since the pandemic. Thus, the CWS has provided promotions such as free day passes and free first-month memberships. This strategy was in stark contrast with the negative reviews regarding membership cancellation and pause issues. In fact, the community manager interviewed said that they did not receive any [direct?] complaints about membership retention.

Complaints are received via email and have not been collected separately in a specific software although they have software for membership management. The primary reason was that most facility-related complaints were reported by the community manager and managed by the FM department in the building. Otherwise, service-related issues such as free refreshment and event preparations are dealt with by the community manager. She also mentioned that the reorganized clustering results are helpful in understanding general views of CWS users as well as comparing their physical settings and services with the results to improve quality. In particular, virtual conferencing and meeting systems with a private space is highly considered by their users.

Second, the facility manager of the ATDC building pointed out that maintenance request data inputs have been an issue for more than 20 years. They expressed that the automated metadata inputs would be very useful because an incorrect category input caused not only difficulty in analyzing the data but also confusion and was time-consuming at the maintenance issue site. Incorrect information disguised real problems by making FM staff spend time identifying irrelevant issues; this issue occurred primarily when a non-professional building user made the category input. Thus, the FM department began controlling the authorization to input category information and to request maintenance to avoid such situations. In this vein, an automated classification system is useful in practice.

Lastly, in the interview with the GTFM manager, an interesting maintenance behavior arose due to COVID-19. Both occupants and FM staff did not want to encounter each other when conducting repairs or maintenance of a facility after COVID-19. For instance, maintenance activities were requested and scheduled when building users were not onsite at the office and space. This led to increases in the time allocated for FM staff to respond to the issues according to the requested times and locations. This implies that the requested times should be recorded and managed as a separate column in the maintenance dataset for efficient staff and resource allocations.

The interviews with the professionals provided multifaceted views that validate the practical contributions of this research. In addition, the interview implied future research to increase CWS user satisfaction by offering the highly ranked preferred facilities and services and reducing the time for scheduling maintenance activities.

CHAPTER 7. CONCLUSION

In this research, the preferences of CWS are investigated and illustrated through the thematic categories of CWS users' experiences and the mixed methods including content analysis, text classification, and clustering. The findings of the analyses highlight the importance of hospitality and technology. The hospitality of community managers is the most important factor for user experience, and technologies, including fast wireless Internet and private and properly equipped spaces, are critical to support video conferencing in the era of hybrid workplaces. A significant difference between the cities in this study does not present on the six-category level. In future research, investigating differences among cities on element and sub-category levels is suggested.

Qualified social media data helped overcome the constraints in collecting data during COVID-19 during which many CWS temporarily closed and CWS users did not show up. The mixed-method approach triangulates the findings. The clustering method helps understand the trends of the reviews, while content analysis provides specific insight and characteristics into designing and operating CWSs. The classification results indicate a fairly strong accuracy. However, labeling more data and the imbalanced data issue should be addressed. This study sheds light on the applications of machine learning with NLP and deep learning-based transformer models in physical workplace management research.

Two technical improvements are expected in future research to extract more insights from big data. First, future studies should resolve imbalance issues in the BERT classification with technical approaches such as MLSMOTE. Second, sentence similarity analysis utilizing a deep learning-based transformer and data manipulation methods is suggested as future work. This process will reduce the iterative steps in selecting seed words and finding related sentences according to categories. This would illustrate the systematic process for grouping sentences about physical workplace issues in reviews and identify satisfaction levels and reasons for discomfort. Lastly, investigating user preferences during or post-COVID-19 through structured interviews and survey-based data would complement this study's findings as well as facilitate building a theoretical user experience model in CWS.

This research also examined the applicability of data integration of different sources in constructing machine learning text classification models with respect to facility maintenance. Adaptation and integration of relevant big data to a targeted dataset increased the machine learning prediction performance for the small dataset. Furthermore, feature engineering based on domain knowledge helped identify integration feasibility and increased prediction performance of the text classification model both qualitatively and quantitatively. In this research, accumulating public facilities' large, open-source data is suggested to accelerate automated data management systems of single or small facilities. For future research, it is recommended to generate data clusters of building blocks based on the type of buildings, such as office buildings, K-12 school facilities, dormitories, libraries, and government buildings. The accumulated database can be utilized to construct datasets of key phrases and words in maintenance according to building types as well as to explore machine learning applications to automate maintenance processes, particularly for small-scale buildings.

APPENDIX A. APPENDICES

A.1 Study I: Category Labeling Schemes

Six categories included are labeled as follows:

Category	**Sub-category**
Staff (community managers)	Hospitality, work support, community support, maintenance of spaces (e.g., cleaning, organizing, and repairing)
Space	IEQ, space configuration, interior design and atmosphere, facilities
Technology	Internet access, printing service, virtual conferencing system, booking system
Service	Networking and community events, work supportive service (childcare and administrative service), leisure programs, refreshment
Accessibility	Location, parking lots
Membership	Membership types, contract period

Each review can have multiple categories according to the context. For instance, when a review includes a sentence, a phrase, or a word related to staff and space, the review is labeled as Staff and Space with a 1. In case of reviews not included in any category, they do not have any flags to indicate a topic (i.e., it will be marked with all zeros).

- Staff (community managers)
    - <u>Definition:</u>

      This category refers to expressions and impressions related to staff such as how they deal with customers, their attitude and support.
    - <u>Example words/ phrases:</u>

      Community managers, managers, staff names, customer service, friendly staff, help, support, great staff, accommodate, etc.
    - <u>Example comments:</u>

      "All of my interactions with the staff were super friendly. They were a welcoming face in the morning, gave me a tour on the first day and were available to help when we needed additional support."

      "On the Friday of our visit they were hosting a tailgating event to celebrate the start of football season."

      "The staff here seem genuinely interested in your work, and provide help and guidance in all areas."

- Space
    - Indoor Environmental Quality (IEQ):
        - <u>Definition:</u>

          This category refers to indoor environments including thermal comfort (temperature), acoustic comfort (noise), visual comfort (lighting), visual/ acoustic privacy, spatial layouts, furniture configurations.

- Example words/phrases:

  Spacious, noisy, quiet, natural light, lighting, layout, furniture, desk, table, chair, etc.

- Example comments:

  "The space itself offers a variety of seating including open space with bar seating, long tables, personal desk chairs, semi-private booths and a variety of sizes of conference rooms available (for an additional cost)."

  "The lighting is not too bright, but enough so that your eyes don't have to squint while working. And compared to other coworking spaces, Strongbox West is, undoubtedly, the quietest. Not a lot of people, but lots of work being done."

• Interior design and atmosphere

- Definition:

  This category ranges from the atmosphere of the space to interior decoration and color schemes.

- Example words/phrases:

  Décor, vibe, atmosphere, colorful art hung, painting, wall, culture, amenities, modern, etc.

- *Example comments:*

  "They have a bevy of incredible, colorful local art hung and painted throughout the space."

  "The decor of the space is trendy industrial but not too distracting to take away from your work. I really liked the illuminated lettering of "Buckhead" in one of the seating areas."

"The decor is clean, warm and welcoming."

- Facilities
  - Definition:

    This category refers to the condition and availability of facilities and amenities in coworking spaces such as 1) Size and availability of conference rooms, meeting rooms and spaces, phone booths, storages, appliances, 2) cleanliness of spaces, bathrooms, and dining spaces, and 3) door locks.
  - Example words/phrases:

    Many meeting rooms, conference rooms, microwave, bathrooms, door lock, phone booth, phone booth
  - *Example comments:*

    "There are also nice bathrooms."

    "The suites have all the amenities that a business owner could ever want or need."

    "Clean and well maintained complex."

- Service
  - Work support
    - Definition:

      This category refers to services and education that support or facilitate tenants' core work such as childcare in a coworking space, work-related education and training, and administrative services.

- Example words/phrases:

  Daycare, baby, toddler, working parents, training, conference, administrative service, virtual office, mail/package delivery, etc.

- Example comments:

  "I definitely recommend this place if you need software training."

  "I attended an event here for a conference."

  "*Administrative Services for when you need them- ease of doing business because of our onsite Ricoh copier, FedEx, USPS postage, and overnight Staples delivery- ease of moving*"

- Events (networking and collaboration)

  - Definition:

    This category refers to community, environments and events that facilitate communication and interaction opportunities between coworking space users as well as with non-members. This category covers hosting of events by coworking space staff and members and the composition of members (diverse, professional, and perceived as desirable).

  - Example words/ phrases:

    Networking event(s), hosting event(s), collaborative environments, collaboration, meet people, friends, entrepreneurs, create a community, community

  - Example comments:

"Staff members were wearing their alma mater's jersey while they invited guests/members to take a break from the grind and enjoy a BBQ lunch, play a game of cornhole and meet other members."

"The tenants are cool. Lots of interesting companies taking off from here."

"I came here for a networking event and was blown away by the beauty of this event space"

"I met interesting entrepreneurs and made good friends there."

- Leisure
  - <u>Definition:</u>

    This category includes leisure and wellness activities at the coworking spaces.
  - <u>Example words/ phrases:</u>

    Yoga, ping-pong, pilates, cardio programs, massages, etc.
  - <u>Example comments:</u>

    "The office events we have are so much fun! We usually have at least a popcorn/beer day, yogurt day, or both during the week. We also have had numerous other events like cookie decorating, manicures, massages, margarita making, etc."
  - *"They're also very good about planning activities to break up the day* - chair *massages, weekly happy hours, quiche and yogurt parfait bars, etc."*
- Refreshment
  - <u>Definition:</u>

    This category includes refreshment-related comments such as food, drinks, coffee, beer, snack, etc.

- Example words/phrases:

  Free coffee, free beer, free snacks, café, food, drinks, barista, etc.

- Example Comments:

  "Although the daily croissants and pastries are detrimental to my waistline, there are always tasty (and healthy) snacks to graze on."

  "They even sell food and drinks."

  "They have coffee, tea (which I think are free), and also Korean snacks and drinks (I think not free)."

• Pets

- Definition:

  This category refers to coworking spaces that have a pet-friendly policy and allow coworking space users to bring their pets.

- Example words/phrases:

  Dog friendly, cat friendly, etc.

- Example comments:

  "Dog friendly!"

  "It was dog friendly, which was great because we wanted to use our dogs in our ceremony."

- Technology

    - <u>Definition:</u>

This category refers to technology to support work and productivity.

    - <u>Example words/ phrases:</u>

Internet, WiFi, printers, monitors, displays, projection screens.

    - <u>Example comments:</u>

"The WiFi is great."

"The internet was sustainable, but I felt the connection could be a bit stronger as I was frequently disconnected."

"They have projection screens, large monitors for briefing and much more!"

- Accessibility

    - <u>Definition</u>*:*

This category includes proximity to public transportation, the availability of parking lots, and commuting distance.

    - <u>Example words/phrases:</u>

Station, parking lots, parking deck, location, accessibility, etc.

    - <u>Example comments:</u>

"Parking exists both in the front and back of the building, although, the front couldn't have been more packed."

"This Ponce de Leon location is SO conveniently located in the city."

"There is a parking deck you can utilize if you're driving and although it's shared parking with a handful of restaurants it didn't seem to be a challenge to get a space."

"I took Marta on my first visit which was super convenient for me since I was coming from Reynoldstown. "

- Membership
  - Definition:

This category is relevant to price and membership contracts.

  - Example words/ phrases:

Price, membership, day pass, monthly contract, charge, reasonable rates, etc.

  - Example Comments:

"I think the price is fair."

"They tried to charge me an extra month even when I notified them in writing days in advance."

A.2 Study I: Hyper-parameters

Support Vector Machine (SVM)

- Kernel
  - Radial Basis Function (RBF)
- Hyperparameters
  - Penalty parameter (C) = 100
  - RBF-related parameter (r) = 0.01

Multi-layer Perceptron (MLP)

- Optimizer: Limited-memory BFGS
- Learning rate: 1×10^{-5}
- Other parameters: Defaults provided by the Sklearn library

Logistic Regression

- Optimizer: Stochastic Average Gradient (SAG) descent
- Regularization: L2
- Other parameters: Defaults provided by the Sklearn library

Bidirectional Encoder Representations (BERT)

- Optimizer: AdamW
- Learning rate: 2×10^{-5}
- Adam Epsilon: 1×10^{-8}
- Epochs: 5
- Others parameters: Defaults provided by the Hugging Face transformers library

A.3 Study II: Confusion matrices

Electrical

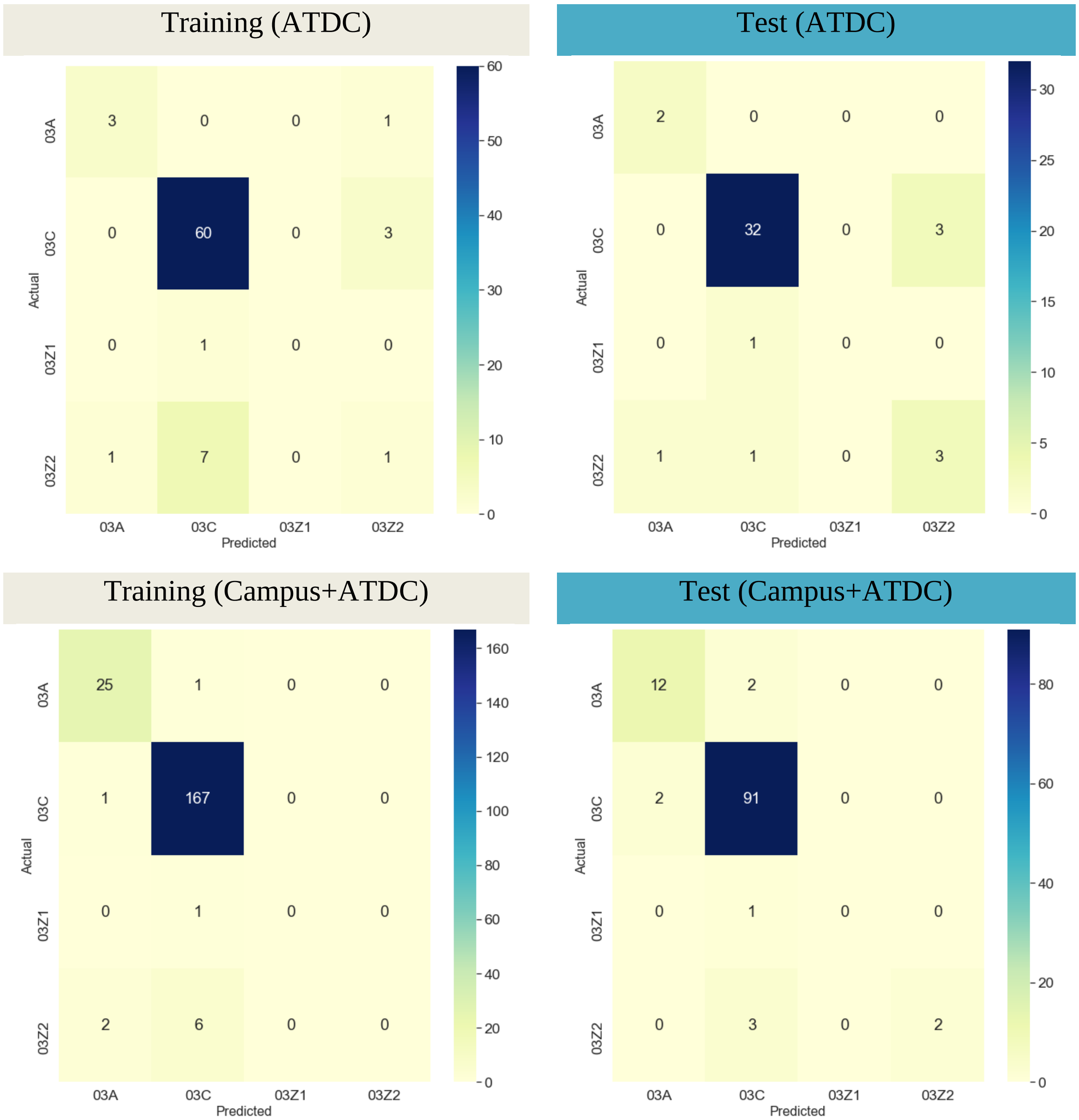

Figure 15. Confusion matrices of the text classifications in electrical

Plumbing

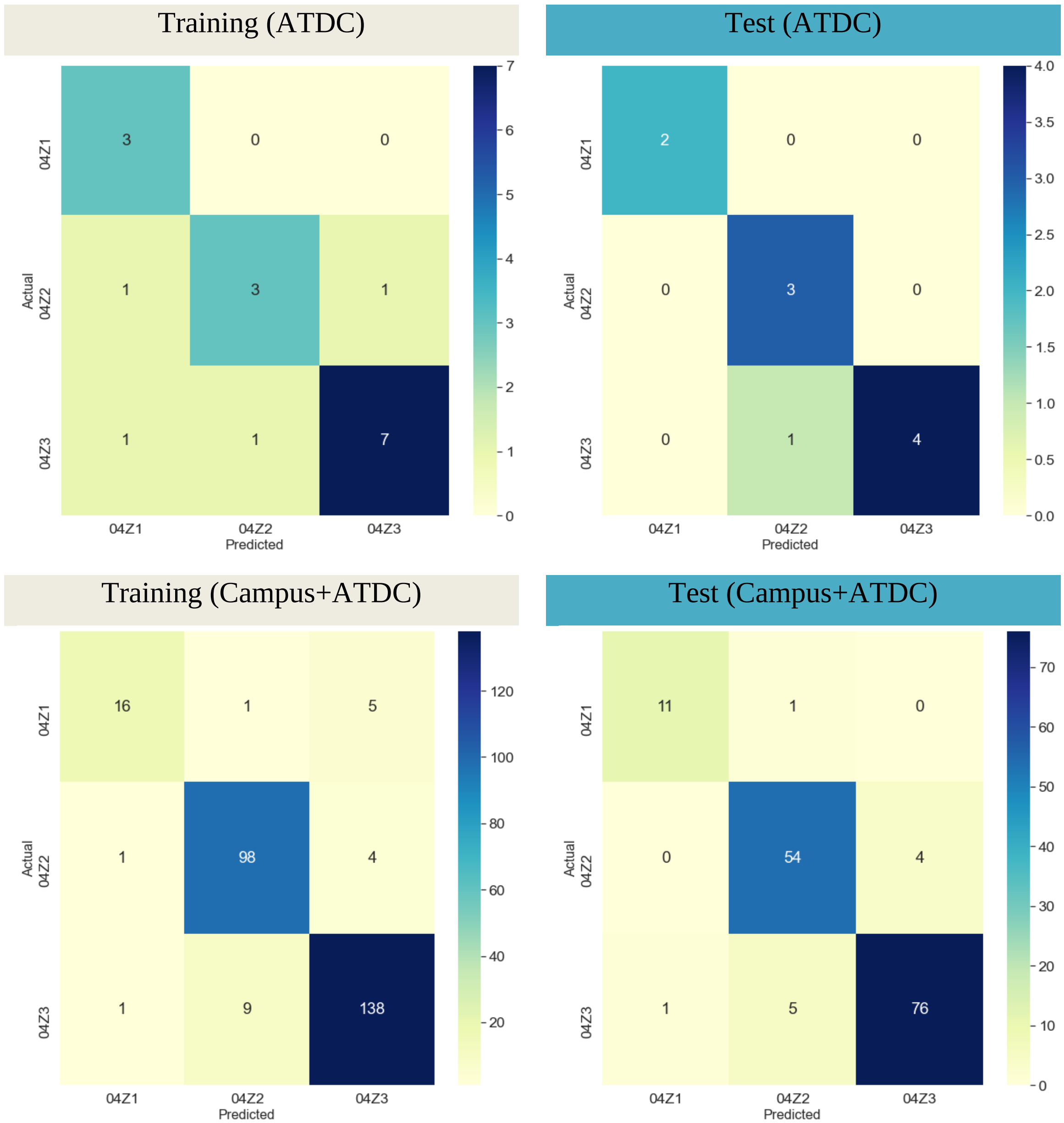

Figure 16. Confusion matrices of the text classifications in plumbing

REFERENCES

Alaei, Ali Reza, Susanne Becken, and Bela Stantic. 2019. “Sentiment Analysis in Tourism: Capitalizing on Big Data.” *Journal of Travel Research* 58(2):175–91. doi: 10.1177/0047287517747753.

Alexander, Keith, Brian Atkin, Jan Bröchner, and Tore Haugen. 2004. *Facilities Management: Innovation and Performance*. Routledge.

Allen, Thomas, and Gunter Henn. 2007. *The Organization and Architecture of Innovation*. Routledge.

Allen, Thomas J. 1970. “Communication Networks in R & D Laboratories.” *R&D Management* 1(1):14–21. doi: 10.1111/j.1467-9310.1970.tb01193.x.

Allen, Thomas J. 1977. “Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R and D Organization.” doi: 10.5860/crl_39_02_139.

Alpaydin, Ethem. 2020. *Introduction to Machine Learning*. MIT press.

Altrabsheh, Nabeela, Mohamed Medhat Gaber, and Mihaela Cocea. 2013. “SA-E: Sentiment Analysis for Education.” Pp. 353–62 in *International conference on intelligent decision technologies*. Vol. 255.

Appel-Meulenbroek, Rianne, Minou Weijs-Perrée, Marko Orel, Felix Gauger, and Andreas Pfnür. 2021. “User Preferences for Coworking Spaces; a Comparison between the

Netherlands, Germany and the Czech Republic." *Review of Managerial Science* 15(7):2025–48. doi: https://doi.org/10.1007/s11846-020-00414-z.

Araszkiewicz, Krystyna. 2017. "Digital Technologies in Facility Management – The State of Practice and Research Challenges." *Procedia Engineering* 196:1034–42. doi: https://doi.org/10.1016/j.proeng.2017.08.059.

Atkin, Brian, and Adrian Brooks. 2015. *Total Facility Management*. John Wiley & Sons.

Balakrishna, Sivadi, M. Thirumaran, and Vijender Kumar Solanki. 2020. "IoT Sensor Data Integration in Healthcare Using Semantics and Machine Learning Approaches." Pp. 275–300 in *A handbook of internet of things in biomedical and cyber physical system*. Springer, Cham.

Bartels, Jos, Oscar Peters, Menno De Jong, Ad Pruyn, and Marjolijn van der Molen. 2010. "Horizontal and Vertical Communication as Determinants of Professional and Organisational Identification." *Personnel Review*. doi: 10.1108/00483481011017426.

Batrinca, Bogdan, and Philip C. Treleaven. 2014. "Social Media Analytics: A Survey of Techniques, Tools and Platforms." *AI and Society* 30(1):89–116. doi: 10.1007/s00146-014-0549-4.

De Been, Iris, and Marion Beijer. 2014. "The Influence of Office Type on Satisfaction and Perceived Productivity Support." *Journal of Facilities Management* 12(2):142–57. doi: 10.1108/JFM-02-2013-0011.

Bernstein, Ethan S., and Stephen Turban. 2018. "The Impact of the 'Open' Workspace on

Human Collaboration." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1753):20170239.

Berrar, Daniel, Philippe Lopes, and Werner Dubitzky. 2019. "Incorporating Domain Knowledge in Machine Learning for Soccer Outcome Prediction." *Machine Learning* 108(1):97–126. doi: 10.1007/s10994-018-5747-8.

Berwald, Sarah. 2008. "From CAD to BIM: The Experience of Architectural Education with Building Information Modeling." Pp. 1–5 in *AEI 2008: Building Integration Solutions*.

Bhatt, Aashutosh, Ankit Patel, Harsh Chheda, and Kiran Gawande. 2015. "Amazon Review Classification and Sentiment Analysis." *International Journal of Computer Science and Information Technologies* 6(6):5107–10.

Bick, Alexander, Adam Blandin, and Karel Mertens. 2020. "Work from Home after the COVID-19 Outbreak." doi: 10.24149/wp2017.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. " O'Reilly Media, Inc."

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.

van Blokland, Anette B. 2018. "London's Hottest Coworking Spaces Have Childcare For Your Kids." *Forbes*. Retrieved December 21, 2020 (https://www.forbes.com/sites/anettebvanblokland/2018/07/17/londons-hottest-coworking-spaces-have-childcare-for-your-kids/?sh=6d2d53167631).

Boland, Brodie, Aaron De Smet, Rob Palter, and Aditya Sanghvi. 2020. “Reimagining the Office and Work Life after COVID-19.”

van Bommel, WJM, and GJ van den Beld. 2004. “Lighting for Work: A Review of Visual and Biological Effects.” *Lighting Research and Technology* 36(4):255–66. doi: 10.1191/1365782804li122oa.

van Bommel, Wout J. M. 2006. “Non-Visual Biological Effect of Lighting and the Practical Meaning for Lighting for Work.” *Applied Ergonomics* 37(4 SPEC. ISS.):461–66. doi: 10.1016/j.apergo.2006.04.009.

Bortolini, Rafaela, and Núria Forcada. 2020. “Analysis of Building Maintenance Requests Using a Text Mining Approach: Building Services Evaluation.” *Building Research & Information* 48(2):207–17. doi: https://doi.org/10.1080/09613218.2019.1609291.

Bouncken, Ricarda B., and Andreas J. Reuschl. 2018. “Coworking-Spaces: How a Phenomenon of the Sharing Economy Builds a Novel Trend for the Workplace and for Entrepreneurship.” *Review of Managerial Science* 12(1):317–34. doi: https://doi.org/10.1007/s11846-016-0215-y.

Bouncken, Ricarda, Martin Ratzmann, Roman Barwinski, and Sascha Kraus. 2020. “Coworking Spaces: Empowerment for Entrepreneurship and Innovation in the Digital and Sharing Economy.” *Journal of Business Research* 114:102–10. doi: https://doi.org/10.1016/j.jbusres.2020.03.033.

Brown, Julie. 2017. “Curating the ‘Third Place’? Coworking and the Mediation of Creativity.” *Geoforum* 82:112–26. doi:

https://doi.org/10.1016/j.geoforum.2017.04.006.

Brownlee, Jason. 2018. *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery.

Button, Annie. 2019. “The Overlooked Attributes of Every Successful Coworking Space.” *Coworker*. Retrieved December 21, 2020 (https://www.coworker.com/mag/the-overlooked-attributes-of-every-successful-coworking-space).

Cabral, Victor, and Willem Van Winden. 2016. “Coworking: An Analysis of Coworking Strategies for Interaction and Innovation.” *International Journal of Knowledge-Based Development* 7(4):357–77. doi: 10.1504/ijkbd.2016.10001777.

Campello, Ricardo J. G. B., Davoud Moulavi, and Jörg Sander. 2013. “Density-Based Clustering Based on Hierarchical Density Estimates.” Pp. 160–72 in *Pacific-Asia conference on knowledge discovery and data mining*. Berlin, Heidelberg: Springer.

Caraher, Lee. 2016. *Millennials & Management: The Essential Guide to Making It Work at Work*. Routledge.

Carlopio, James R. 1996. “Construct Validity of a Physical Work Environment Satisfaction Questionnaire.” *Journal of Occupational Health Psychology* 1(3):330. doi: 10.1037/1076-8998.1.3.330.

Chang, Chen-Yen, and Ping-Kun Chen. 2005. “Human Response to Window Views and Indoor Plants in the Workplace.” *HortScience* 40(5):1354–59. doi: 10.21273/hortsci.40.5.1354.

Charte, Francisco, Antonio J. Rivera, María J. del Jesus, and Francisco Herrera. 2015. “MLSMOTE: Approaching Imbalanced Multilabel Learning through Synthetic Instance Generation.” *Knowledge-Based Systems* 89:385–97. doi: https://doi.org/10.1016/j.knosys.2015.07.019.

Chinazzo, Giorgia. 2021. “Investigating the Indoor Environmental Quality of Different Workplaces through Web-Scraping and Text-Mining of Glassdoor Reviews.” *Building Research & Information* 1–19. doi: https://doi.org/10.1080/09613218.2021.1908879.

Clements-Croome, D. 2006. *Creating the Productive Workplace.* 2nd ed. London: Taylor & Francis.

Cortes, Corinna, and Vladimir Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20(3):273–97. doi: 10.1007/bf00994018.

Cronin, Robert M., Daniel Fabbri, Joshua C. Denny, S. Trent Rosenbloom, and Gretchen Purcell Jackson. 2017. “A Comparison of Rule-Based and Machine Learning Approaches for Classifying Patient Portal Messages.” *International Journal of Medical Informatics* 105:110–20. doi: 10.1016/j.ijmedinf.2017.06.004.

Davenport, Thomas H. 2005. *Thinking for a Living: How to Get Better Performances and Results from Knowledge Workers*. Harvard Business Press.

Dearden, Lorraine, Howard Reed, and John Van Reenen. 2006. “The Impact of Training on Productivity and Wages: Evidence from British Panel Data.” *Oxford Bulletin of Economics and Statistics* 68(4):397–421. doi: 10.1920/wp.ifs.2005.0516.

DeFilippis, Evan, Stephen Michael Impink, Madison Singell, Jeffrey T. Polzer, and Raffaella Sadun. 2020. “Collaborating during Coronavirus: The Impact of COVID-19 on the Nature of Work.” *NBER Working Paper* (w27612). doi: 10.3386/w27612.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *ArXiv Preprint ArXiv:1810.04805*. doi: 10.48550/arXiv.1810.04805.

Dong, Xin Luna, and Theodoros Rekatsinas. 2018. “Data Integration and Machine Learning: A Natural Synergy.” Pp. 1645–1650 in *Proceedings of the 2018 International Conference on Management of Data*, *SIGMOD ’18*. New York, NY, USA: Association for Computing Machinery.

Duan, Wenjing, Yang Yu, Qing Cao, and Stuart Levy. 2016. “Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach.” *Cornell Hospitality Quarterly* 57(3):282–96. doi: 10.1177/1938965515620483.

Endrissat, Nada, and Aurélie Leclercq-Vandelannoitte. 2021. “From Sites to Vibes: Technology and the Spatial Production of Coworking Spaces.” *Information and Organization* 31(4):100353. doi: https://doi.org/10.1016/j.infoandorg.2021.100353.

Faculty of Arts and Sciences Columbia University. 2022. “Working from Home: WiFi/Network Considerations.” *Columbia University*. Retrieved February 27, 2022 (https://fas.columbia.edu/home/contingency-planning/working-home-wifinetwork-considerations).

Fang, ZIGENG, MICHAEL Pitt, and SEAN Hanna. 2019. “Machine Learning in Facilities

& Asset Management.” in *Proceedings of the 25th annual Pacific rim real estate society (PRRES) conference*. PRRES.

Feldscher, Karen. 2021. “What Will It Be like When COVID-19 Becomes Endemic?” *Harvard T.H. Chan School of Public Health*. Retrieved April 12, 2022 (https://www.hsph.harvard.edu/news/features/what-will-it-be-like-when-covid-19-becomes-endemic/).

Ferreira, Edmund John, Annatjie W. Erasmus, and Darelle Groenewald. 2010. *Administrative Management*. Juta and Company Ltd.

Floyd, Myron F., John O. Spengler, Jason E. Maddock, Paul H. Gobster, and Luis J. Suau. 2008. “Park-Based Physical Activity in Diverse Communities of Two US Cities: An Observational Study.” *American Journal of Preventive Medicine* 34(4):299–305. doi: 10.1016/j.amepre.2008.01.009.

Foertsch, Carsten. 2011. “The Coworker’s Profile.” *Deskmag*. Retrieved December 21, 2020 (https://www.deskmag.com/en/coworkers/the-coworkers-global-coworking-survey-168).

Fradkin, Andrey, Elena Grewal, Dave Holtz, and Matthew Pearson. 2015. “Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb.” *EC* 15:15–19. doi: 10.1145/2764468.2764528.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics New York.

Garrett, Lyndon E., Gretchen M. Spreitzer, and Peter A. Bacevice. 2017. “Co-Constructing a Sense of Community at Work: The Emergence of Community in Coworking Spaces.” *Organization Studies* 38(6):821–42. doi: 10.1177/0170840616685354.

Gilbert, C. H. E., and Erric Hutto. 2014. “Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” P. 82 in *Eighth International Conference on Weblogs and Social Media (ICWSM-14).* Vol. 81.

Gliedt, Travis, and Christina E. Hoicka. 2015. “Energy Upgrades as Financial or Strategic Investment? Energy Star Property Owners and Managers Improving Building Energy Performance.” *Applied Energy* 147:430–43.

Goldman Sachs. 2019. “Millennials Commin g of Age.” *Goldman Sachs*. Retrieved March 2, 2020 (https://www.goldmansachs.com/insights/archive/millennials/).

Greenberger, Ellen, Wendy A. Goldberg, Sharon Hamill, Robin O’Neil, and Constance K. Payne. 1989. “Contributions of a Supportive Work Environment to Parents’ Well-Being and Orientation to Work.” *American Journal of Community Psychology* 17(6):755–83. doi: 10.1007/bf00922737.

Gunawan, Michael Wu. 2018. “Workplace Wellness and Active Lifestyle in Coworking Spaces.” *SpaceAsia*. Retrieved December 22, 2020 (https://blog.flyspaces.com/incorporating-wellness-and-active-lifestyle-in-coworking-spaces#1).

Gunay, H. Burak, Weiming Shen, and Chunsheng Yang. 2019. “Text-Mining Building Maintenance Work Orders for Component Fault Frequency.” *Building Research &*

Information 47(5):518–33. doi: https://doi.org/10.1080/09613218.2018.1459004.

Han, Hye-Sun. 2013. “A Study on the Composition and Characteristic of Coworking Space.” *Korean Institute of Interior Design Journal* 22(5):276–83. doi: 10.14774/jkiid.2013.22.5.276.

Han, Jiawei, Micheline Kamber, and Jian Pei. 2012. “Getting to Know Your Data.” Pp. 39–82 in *Data Mining: Concepts and techniques*. Vol. 3. Burlington, MA: Morgan Kaufmann.

Hassanain, Mohammad A., Ali K. Alnuaimi, and Muizz O. Sanni-Anibire. 2018. “Post Occupancy Evaluation of a Flexible Workplace Facility in Saudi Arabia.” *Journal of Facilities Management* 16(2):102–18. doi: 10.1108/JFM-05-2017-0021.

Von Hippel, Eric. 2001. “Learning from Open-Source Software.” *MIT Sloan Management Review* 42(4):82–86. doi: 10.2139/ssrn.290305.

Ho, Tin Kam. 1995. “Random Decision Forests.” Pp. 278–82 in *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE.

Hoendervanger, Jan Gerard, Nico W. Van Yperen, Mark P. Mobach, and Casper J. Albers. 2019. “Perceived Fit in Activity-Based Work Environments and Its Impact on Satisfaction and Performance.” *Journal of Environmental Psychology* 65:101339. doi: 10.1016/j.jenvp.2019.101339.

Hong, Sungil, Junghyun Kim, and Eunhwa Yang. 2022. “Automated Text Classification of Historical Maintenance Data of Higher Education Buildings Using Text Mining and Machine Learning Methods." *Journal of Architectural Engineering* 28(1). doi: https://doi.org/10.1061/(ASCE)AE.1943-5568.0000522.

Horizons, Bright. 2004. "The Real Savings from Employer Sponsored Child Care: Investment Impact Study Results." *Watertown, Mass.: Bright Horizons Family Solutions*.

Al Horr, Yousef, Mohammed Arif, Amit Kaushik, Ahmed Mazroei, Martha Katafygiotou, and Esam Elsarrag. 2016. "Occupant Productivity and Office Indoor Environment Quality: A Review of the Literature." *Building and Environment* 105:369–89. doi: 10.1016/j.buildenv.2016.06.001.

Huang, Li, Yingxin Zhu, Qin Ouyang, and Bin Cao. 2012. "A Study on the Effects of Thermal, Luminous, and Acoustic Environments on Indoor Environmental Comfort in Offices." *Building and Environment* 49(1):304–9. doi: 10.1016/j.buildenv.2011.07.022.

Huang, Ming-Hui, and Roland T. Rust. 2018. "Artificial Intelligence in Service." *Journal of Service Research* 21(2):155–72. doi: 10.1177/1094670517752459.

Huang, Yifu. 2021. "The Role of Community Manager in a Coworking Space." *Coworking Resources*. Retrieved April 11, 2022 (https://www.coworkingresources.org/blog/community-manager).

International WELL Building Institute. 2022. "WELL Certified." *International WELL Building Institute*. Retrieved March 20, 2022 (https://www.wellcertified.com/).

Jain, Ruchi, and Surinder Kaur. 2014. “Impact of Work Environment on Job Satisfaction.” *International Journal of Scientific and Research Publications* 4(1):1–8.

Jasiński, Artur. 2020. “Public Space or Safe Space–Remarks during the COVID-19 Pandemic.” *Technical Transactions* 117(1). doi: 10.37705/techtrans/e2020020.

Joachims, Thorsten. 1998. “Text Categorization with Support Vector Machines: Learning with Many Relevant Features.” Pp. 137–42 in *European conference on machine learning*. Springer.

Johnson, Cat. 2018. “Inside The World’s First Coworking Space With Full Childcare: A Q&A With Third Door’s Shazia Mustafa | AllWork.Space.” *Allwork.* Retrieved December 21, 2020 (https://allwork.space/2018/12/inside-the-worlds-first-coworking-space-with-full-childcare-a-qa-with-third-doors-shazia-mustafa/).

Kaya, NAz, and Helen Epps. 2004. “Relationship between Color and Emotion: A Study of College Students.” *College Student Journal* 38(3):396.

Kharde, Vishal, and Prof Sonawane. 2016. “Sentiment Analysis of Twitter Data: A Survey of Techniques.” *International Journal of Computer Applications* 139(11):5–15. doi: 10.5120/ijca2016908625.

Kim, Jungsoo, and Richard de Dear. 2012. “Nonlinear Relationships between Individual IEQ Factors and Overall Workspace Satisfaction.” *Building and Environment* 49(1):33–40. doi: 10.1016/j.buildenv.2011.09.022.

Kim, Jungsoo, and Richard de Dear. 2013. “Workspace Satisfaction: The Privacy-

Communication Trade-off in Open-Plan Offices." *Journal of Environmental Psychology* 36:18–26. doi: 10.1016/j.jenvp.2013.06.007.

Kleppner, Daniel. 2010. "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age."

Kojo, Ilona Inka Vuokko, and Suvi Nenonen. 2014. "User Experience in an Academic Coworking Place: The Case of Aalto University's Design Factory." Pp. 341–53 in *Cib Facilieties Management Conference Using Facilities in and Open World - Creating Value for all Stakeholders. Joint CIB W070, W111 and W118 Conferece, Technical University of Denmark, Copenhagen 21-*, edited by P. A. Jensen.

Kojo, Ilona Inka Vuokko, and Suvi Nenonen. 2015. "Places for Multi - Locational Work– Opportunities for Facilities Management." *Facilities* 33:20–37. doi: 10.1108/F-05-2013-0043.

Kojo, Ilona Inka Vuokko, and Suvi Nenonen. 2017. "Evolution of Co-Working Places: Drivers and Possibilities." *Intelligent Buildings International* 9(3):164–75. doi: https://doi.org/10.1080/17508975.2014.987640.

Krueger, Alan, and Cecilia Rouse. 1998. "The Effect of Workplace Education on Earnings, Turnover, and Job Performance." *Journal of Labor Economics* 16(1):61–94. doi: 10.1086/209882.

Kumar, Vinnet, Avishek Lahiri, and Orhan Bahadir Dogan. 2018. "A Strategic Framework for a Profitable Business Model in the Sharing Economy." *Industrial Marketing Management* 69:147–60. doi: 10.1016/j.indmarman.2017.08.021.

Kwallek, Nancy, C. M. Lewis, J. W. D. Lin-Hsiao, and H. Woodson. 1996. "Effects of Nine Monochromatic Office Interior Colors on Clerical Tasks and Worker Mood." *Color Research & Application* 21(6):448–58. doi: 10.1002/(SICI)1520-6378(199612)21:6<448::AID-COL7>3.0.CO;2-W.

Lait, Jana, and Jean Wallace. 2002. "Stress at Work: A Study of Organizational-Professional Conflict and Unmet Expectations." *Relations Industrielles/Industrial Relations* 57(3):463–90. doi: 10.7202/006886ar.

Lamba, Manika, and Margam Madhusudhan. 2018. "Application of Sentiment Analysis in Libraries to Provide Temporal Information Service: A Case Study on Various Facets of Productivity." *Social Network Analysis and Mining* 8(1):63. doi: 10.1007/s13278-018-0541-y.

Lee, So Young, and Jay L. Brand. 2005. "Effects of Control over Office Workspace on Perceptions of the Work Environment and Work Outcomes." *Journal of Environmental Psychology* 25(3):323–33. doi: 10.1016/j.jenvp.2005.08.001.

Lee, Sohyun Sarah. 2018. "Third Places to Work in the Digital Age: Implications from Coworking Space Users' Motivations and Preferred Environmental Features." doi: 10.7298/X4251GD5.

Liddy, Elizabeth D. 2001. "Natural Language Processing."

Liegl, Michael. 2014. "Nomadicity and the Care of Place—on the Aesthetic and Affective Organization of Space in Freelance Creative Work." *Computer Supported Cooperative Work (CSCW)* 23(2):163–83. doi: 10.1007/s10606-014-9198-x.

Liimatainen, Karoliina. 2015. “Supporting Inter-Organizational Collaboration in Coworking Clusters: The Role of Place, Community and Coordination.”

Lopez, Russell P., and H. Patricia Hynes. 2006. “Obesity, Physical Activity, and the Urban Environment: Public Health Research Needs.” *Environmental Health* 5(1):25. doi: 10.1186/1476-069x-5-25.

Van Der Maaten, Laurens, Eric Postma, and Jaap Van den Herik. 2009. “Dimensionality Reduction: A Comparative.” *Journal of Machine Learning Research* 10(66–71):13.

MacMillan, Kyle, Tarun Mangla, James Saxon, and Nick Feamster. 2021. “Measuring the Performance and Network Utilization of Popular Video Conferencing Applications.” *ArXiv Preprint ArXiv:2105.13478*. doi: 10.1145/3487552.3487842.

Malik, Sumaira H., Holly Blake, and L. Suzanne Suggs. 2014. “A Systematic Review of Workplace Health Promotion Interventions for Increasing Physical Activity.” *British Journal of Health Psychology* 19(1):149–80. doi: 10.1111/bjhp.12052.

Mariotti, Ilaria, Carolina Pacchi, and Stefano Di Vita. 2017. “Co-Working Spaces in Milan: Location Patterns and Urban Effects.” *Journal of Urban Technology* 24(3):47–66. doi: 10.1080/10630732.2017.1311556.

Marquardt, Clinton J. G., Jennifer Ann Veitch, and Kate E. Charles. 2002. *Environmental Satisfaction with Open-Plan Office Furniture Design and Layout*. Ottawa: Citeseer.

Mas, Alexandre, and Amanda Pallais. 2020. “Alternative Work Arrangements.” *Annual Review of Economics* 12:631-658. doi: https://doi.org/10.1146/annurev-economics-

022020-032512.

Mayerhoffer, Manuel. 2021. "The Impact of Covid-19 on Coworking Spaces: Evidence from Germany." *Journal of Corporate Real Estate* 23(3):170–85. doi: 10.1108/JCRE-10-2020-0044.

McArthur, J. J., and Colin Powell. 2020. "Health and Wellness in Commercial Buildings: Systematic Review of Sustainable Building Rating Systems and Alignment with Contemporary Research." *Building and Environment* 171:106635. doi: 10.1016/j.buildenv.2019.106635.

McInnes, Leland, John Healy, and Steve Astels. 2017. "Hdbscan: Hierarchical Density Based Clustering." *Journal of Open Source Software* 2(11):205. doi: 10.21105/joss.00205.

McInnes, Leland, John Healy, and James Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." doi: 10.21105/joss.00861.

Merkel, Janet. 2015. "Coworking in the City." *Ephemera* 15(2):121–39. doi: https://openaccess.city.ac.uk/id/eprint/14478.

Meyers-Levy, Joan, and Rui Zhu. 2007. "The Influence of Ceiling Height: The Effect of Priming on the Type of Processing That People Use." *Journal of Consumer Research* 34(2):174–86. doi: 10.1086/519146.

Michener, William K. 2006. "Meta-Information Concepts for Ecological Data Management." *Ecological Informatics* 1(1):3–7. doi:

https://doi.org/10.1016/j.ecoinf.2005.08.004.

Mo, Yunjeong, Dong Zhao, Jing Du, Matt Syal, Azizan Aziz, and Heng Li. 2020. "Automated Staff Assignment for Building Maintenance Using Natural Language Processing." *Automation in Construction* 113:103150. doi: https://doi.org/10.1016/j.autcon.2020.103150.

Moezzi, Mithra, and John Goins. 2011. "Text Mining for Occupant Perspectives on the Physical Workplace." *Building Research & Information* 39(2):169–82. doi: 10.1080/09613218.2011.556008.

Morisson, Arnault. 2018. "A Typology of Places in the Knowledge Economy: Towards the Fourth Place." Pp. 444–51 in *International symposium on new metropolitan perspectives*. Springer.

Nembhard, Melanie D., D. Jeff Burton, and Joel M. Cohen. 2020. "Ventilation Use in Nonmedical Settings during COVID-19: Cleaning Protocol, Maintenance, and Recommendations." *Toxicology and Industrial Health* 36(9):644–53. doi: https://doi.org/10.1177/0748233720967528.

Ng, Cheuk Fan. 2010. "Teleworker's Home Office: An Extension of Corporate Office?" *Facilities* 28(3/4):137–55. doi: doi.org/10.1108/02632771011023113.

Ng, Eddy S. W., Linda Schweitzer, and Sean T. Lyons. 2010. "New Generation, Great Expectations: A Field Study of the Millennial Generation." *Journal of Business and Psychology* 25(2):281–92. doi: 10.1007/s10869-010-9159-4.

Nonaka, Ikujiro. 1994. “A Dynamic Theory of Organizational Knowledge Creation.” *Organization Science* 5(1):14–37. doi: 10.1287/orsc.5.1.14.

Oldenburg, Ramon, and Dennis Brissett. 1982. “The Third Place.” *Qualitative Sociology* 5(4):265–84. doi: 10.1007/BF00986754.

Orel, Marko. 2019. “Supporting Work–Life Balance with the Use of Coworking Spaces.” *Equality, Diversity and Inclusion: An International Journal*. doi: https://doi.org/10.1108/EDI-01-2019-0038.

Orel, Marko, and María del Mar Alonso Almeida. 2019. “The Ambience of Collaboration in Coworking Environments.” *Journal of Corporate Real Estate*. doi: 10.1108/jcre-12-2018-0050.

Payne, Stephanie C., Allison L. Cook, and Ismael Diaz. 2012. “Understanding Childcare Satisfaction and Its Effect on Workplace Outcomes: The Convenience Factor and the Mediating Role of Work-family Conflict.” *Journal of Occupational and Organizational Psychology* 85(2):225–44. doi: 10.1111/j.2044-8325.2011.02026.x.

Petriglieri, Gianpiero, Susan Ashford, and Amy Wrzesniewski. 2018. “Thriving in the Gig Economy.” *HBR’S 10 MUST* 109.

Philander, Kahlil, and Y. Zhong. 2016. “Twitter Sentiment Analysis: Capturing Sentiment from Integrated Resort Tweets.” *International Journal of Hospitality Management* 55(2016):16–24. doi: 10.1016/j.ijhm.2016.02.001.

Potamias, Michalis. 2012. “The Warm-Start Bias of Yelp Ratings.” *ArXiv Preprint*

ArXiv:1202.5713. doi: 10.48550/arXiv.1202.5713.

Prabhakar, Anand M., McKinley Glover, Pamela W. Schaefer, and James A. Brink. 2020. "Academic Radiology Departmental Operational Strategy Related to the Coronavirus Disease 2019 (COVID-19) Pandemic." *Journal of the American College of Radiology* 17(6):730–33. doi: https://doi.org/10.1016/j.jacr.2020.04.004.

Pranckevičius, Tomas, and Virginijus Marcinkevičius. 2017. "Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification." *Baltic Journal of Modern Computing* 5(2):221. doi: https://doi.org/10.22364/bjmc.2017.5.2.05.

Przybyszewski, P., S. Dziewiątkowski, S. Jaszczur, M. Śmiech, and M. Szczuka. 2017. "Use of Domain Knowledge and Feature Engineering in Helping AI to Play Hearthstone." Pp. 143–48 in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*.

Raffaele, Catherine, and Julia Connell. 2016. "Telecommuting and Co-Working Communities: What Are the Implications for Individual and Organizational Flexibility?" Pp. 21–35 in *Flexible work organizations*. Springer.

Rambocas, Meena, and João Gama. 2013. *Marketing Research: The Role of Sentiment Analysis*. Universidade do Porto, Faculdade de Economia do Porto.

Ramos, Juan. 2003. "Using Tf-Idf to Determine Word Relevance in Document Queries." Pp. 133–42 in *Proceedings of the first instructional conference on machine learning*. Vol. 242. Piscataway, NJ.

Rashid, Mahbub, Kevin Kampschroer, Jean Wineman, and Craig Zimring. 2006. "Spatial Layout and Face-to-Face Interaction in Offices—a Study of the Mechanisms of Spatial Effects on Face-to-Face Interaction." *Environment and Planning B: Planning and Design* 33(6):825–44. doi: 10.1068/b31123.

Rebala, Gopinath, Ajay Ravi, and Sanjay Churiwala. 2019. *An Introduction to Machine Learning*. Springer.

Reimers, Nils, and Iryna Gurevych. 2019. "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks." in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

de Reuver, Mark, Carsten Sørensen, and Rahul C. Basole. 2018. "The Digital Platform: A Research Agenda." *Journal of Information Technology* 33(2):124–35. doi: 10.1057/s41265-016-0033-3.

Rish, Irina. 2001. "An Empirical Study of the Naive Bayes Classifier." Pp. 41–46 in *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3.

Rubinstein, Dana, and Nicole Hong. 2022. "As Remote Work Becomes Permanent, Can Manhattan Adapt?" *The New York Times*. Retrieved April 12, 2022 (https://www.nytimes.com/2022/04/11/nyregion/remote-work-hybrid-manhattan.html).

Rumelhart, David E., Bernard Widrow, and Michael A. Lehr. 1994. "The Basic Ideas in Neural Networks." *Communications of the ACM* 37(3):87–93. doi:

10.1145/175247.175256.

Sallis, James F., Myron F. Floyd, Daniel A. Rodríguez, and Brian E. Saelens. 2012. "Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease." *Circulation* 125(5):729–37. doi: 10.1161/circulationaha.110.969022.

Sanborn, Bonnie. 2015. "Building Community through Coworking: A Case Study of Spatial Factors Affecting Member Satisfaction with Coworkspaces and Collaborative Activity." doi: https://hdl.handle.net/1813/40891.

Schapire, Robert E. 1990. "The Strength of Weak Learnability." *Machine Learning* 5(2):197–227. doi: https://doi.org/10.1007/BF00116037.

Schor, Juliet. 2016. "Debating the Sharing Economy." *Journal of Self-Governance and Management Economics* 4(3):7–22. doi: 10.22381/jsme4320161.

Scott, Sam, and Stan Matwin. 1999. "Feature Engineering for Text Classification." Pp. 379–88 in *ICML*. Vol. 99. Citeseer.

Shah, Foram P., and Vibha Patel. 2016. "A Review on Feature Selection and Feature Extraction for Text Classification." Pp. 2264–68 in *2016 international conference on wireless communications, signal processing and networking (WiSPNET)*. IEEE.

Shishegar, N., and M. Boubekri. 2016. "Natural Light and Productivity: Analyzing the Impacts of Daylighting on Students' and Workers' Health and Alertness." Pp. 18–19 in *Proceedings of the International Conference on "Health, Biological and Life Science"(HBLS-16), Istanbul, Turkey*.

Singer-Velush, N., K. Sherman, and E. Anderson. 2020. "Microsoft Analyzed Data on Its Newly Remote Workforce." *Harvard Business Review*. Retrieved (https://hbr.org/2020/07/microsoft-analyzed-data-on-its-newly-remote-workforce).

Skinner, Natalie Jane, and Janine Chapman. 2013. "Work-Life Balance and Family Friendly Policies."

Slack. 2020. "Report: Remote Work in the Age of Covid-19." *Slack*. Retrieved (https://slack.com/blog/collaboration/report-remote-work-during-coronavirus).

Smith, David C., Sandy Romero, Derek Daniels, and Marc Kunesch. 2020. *The Changing Role of Coworking in the Workplace Ecosystem*.

Spinuzzi, Clay. 2012. "Working Alone Together: Coworking as Emergent Collaborative Activity." *Journal of Business and Technical Communication* 26(4):399–441. doi: https://doi.org/10.1177/1050651912444070.

Spreitzer, Gretchen M., Lindsey Cameron, and Lyndon Garrett. 2017. "Alternative Work Arrangements: Two Images of the New World of Work." *Annual Review of Organizational Psychology and Organizational Behavior* 4:473–99. doi: https://doi.org/10.1146/annurev-orgpsych-032516-113332.

Stieg, Cory. 2021. "What Living with Endemic Covid Will Look like in 2022 and Future." *CNBC*. Retrieved April 12, 2022 (https://www.cnbc.com/2021/12/15/what-living-with-endemic-covid-will-look-like-in-2022-and-future.html).

Surman, Tonya. 2013. "Building Social Entrepreneurship through the Power of

Coworking.” *Innovations: Technology, Governance, Globalization* 8(3–4):189–95. doi: 10.1162/inov_a_00195.

Sutherland, Will, and Mohammad Hossein Jarrahi. 2018. “The Sharing Economy and Digital Platforms: A Review and Research Agenda.” *International Journal of Information Management* 43:328–41. doi: 10.1016/j.ijinfomgt.2018.07.004.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. “Data Sharing by Scientists: Practices and Perceptions.” *PloS One* 6(6):e21101. doi: 10.1371/journal.pone.0021101.

Townsend, Anthony M., Samuel M. DeMarie, and Anthony Hendrickson. 2000. “Virtual Teams: Technology and the Workplace of the Future.” *IEEE Engineering Management Review* 28(2):69–80. doi: 10.5465/ame.1998.1109047.

U.S. General Services Administration. 2022. “National O&M and Consolidated Specifications.” *U.S. General Services Administration*. Retrieved April 13, 2022 (https://www.gsa.gov/real-estate/facilities-management/facilities-operations/national-om-and-consolidated-specifications).

Uda, Tadashi. 2013. “What Is Coworking? A Theoretical Study on the Concept of Coworking.” *A Theoretical Study on the Concept of Coworking (December 27, 2013)*. doi: http://dx.doi.org/10.2139/ssrn.2937194.

United States Census Bureau. 2018. “2018 American Community Survey (ACS).” *United States Census Bureau*. Retrieved March 9, 2020 (https://www.census.gov/programs-surveys/acs.html).

Villeneuve, Hannah, and William O'Brien. 2020. "Listen to the Guests: Text-Mining Airbnb Reviews to Explore Indoor Environmental Quality." *Building and Environment* 169:106555. doi: 10.1016/j.buildenv.2019.106555.

Warkentin, Shahrzad. 2020. "Incredible Co-Working Spaces That Also Offer Childcare." *Red Tricycle*. Retrieved December 21, 2020 (https://redtri.com/co-working-spaces-that-offer-child-care/slide/1).

Waters-Lynch, Julian, and Jason Potts. 2017. "The Social Economy of Coworking Spaces: A Focal Point Model of Coordination." *Review of Social Economy* 75(4):417–33. doi: https://doi.org/10.1080/00346764.2016.1269938.

Waters-Lynch, Julian, Jason Potts, Tim Butcher, Jago Dodson, and Joe Hurley. 2016. "Coworking: A Transdisciplinary Overview." *Available at SSRN 2712217*. doi: http://dx.doi.org/10.2139/ssrn.2712217.

Weber, D., and K. E. Hedges. 2008. "From CAD to BIM: The Engineering Student Perspective." Pp. 1–8 in *AEI 2008: Building Integration Solutions*.

Weijs-Perrée, Minou, HAJA Rianne Appel-Meulenbroek, Felix Gauger, Andreas Pfnür, and Marko Orel. 2020. "Differences in User Preferences across European Coworking Spaces." Pp. 12–23 in *2nd Transdisciplinary Workplace Research conference*. TWR network.

Weijs-Perrée, Minou, Rianne Appel-Meulenbroek, Theo Arentze, and Georges Romme.

2019. “The Influence of the Physical Work Environment of Business Centres on Social Networking and Knowledge Sharing in the Netherlands.” *Intelligent Buildings International* 11(2):105–25. doi: https://doi.org/10.1080/17508975.2019.1574705.

Yang, Eunhwa, Catherine Bisson, and Bonnie Eaton Sanborn. 2019. “Coworking Space as a Third-Fourth Place: Changing Models of a Hybrid Space in Corporate Real Estate.” *Journal of Corporate Real Estate*. doi: https://doi.org/10.1108/JCRE-12-2018-0051.

Yang, Eunhwa, Yujin Kim, and Sungil Hong. 2021. “Does Working-from-Home Work?: Experience of Working-from-Home and the New Value of Corporate Offices Post-COVID-19.” *Journal of Corporate Real Estate*. doi: https://doi.org/10.1108/JCRE-04-2021-0015.

Zhai, Yujia, Wei Song, Xianjun Liu, Lizhen Liu, and Xinlei Zhao. 2018. “A Chi-Square Statistics Based Feature Selection Method in Text Classification.” Pp. 160–63 in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*. IEEE.